Atlantis Highlights in Engineering (AHE), volume 1
International Conference on Science and Technology (ICST 2018)

Correspondence Analysis to Identify Alumni Profiles

A Case Study on ITB Alumni Class of 2008

Khumaeroh Dwi Nur'aini[1], Dessy Rizki Suryani[2], Minuk Riyana[3], Murni Sianturi[6]
Mathematics Education Department
Universitas Musamus
Merauke, Indonesia
[1]khumaerohdwinuraini@gmail.com,
[2]drsuryani7@gmail.com, [3]minuk_fkip@unmus.ac.id,
[6]murni@unmus.ac.id

Utriweni Mukhaiyar[4]
Faculty of Mathematics and Natural Sciences
Institut Teknologi Bandung
Bandung, Indonesia
[4]utriweni@math.itb.ac.id

Evy Nurvitasari[5]
Chemistry Education Department
Universitas Musamus
Merauke, Indonesia
[5]evy_fkip@unmus.ac.id

Abstrak—In today's world of work, the rate of competition among job-hunters is getting increasingly higher. In addition, some companies want to recruit human resources who have more work skills and knowledge in order to improve their quality and productivity. Both skills and knowledge are necessary regardless of the type of profession or work field and difficulties that should be conquered at work. Skills and knowledge may be observed in the score indicated in the GPA and programs one followed while she/ he was at college. This research is designed to understand correspondence analysis and apply the method to observe the closeness among variables of ITB alumni class of 2008, namely between the study program and the waiting period to get the first job. Research findings conclude that ITB's Faculty of Industrial Technology has two study programs with the fastest period of waiting to get the first job (which is less than 2 months) while ITB's Faculty of Mathematics and Natural Sciences has two study programs with the longest waiting period (which is more than 12 months). Due to competitiveness at work and among companies, prospective employees who have good skills (which may be indicated by a higher GPA) are absolutely required. This research is expected to provide input for ITB in order to improve quality of its alumni and also to be taken into consideration by prospective college students in selecting which study program and university they will choose.

Keyword— correspondence analysis; tracer study; qualitative variables; principal component; cross tabulation

I. INTRODUCTION

In today's world of work, the level of competition between job seekers is getting increasingly higher. This may due to a large number of job seekers, which, unfortunately, are inversely proportional to the employment level. As a result of today's developments in science and technology, many companies need human resources that have more work skills and knowledge in order to improve the quality and productivity of those companies.

The skills and knowledge needed also depend on the profession or field of business and the level of difficulty an employee will have to deal with. Nowadays, many State-Owned Enterprises (SOEs), private companies, and foreign companies generally require applicants to have a minimum GPA of 3.00 (for graduates of a Bachelor's degree program) and 3.25 (for graduates of a Master's degree program). Even, to take the Prospective Civil Servant test, test-takers are required to have a minimum GPA of 2.70 for state university graduates and 2.90 for private university graduates.

However, it turns out that a GPA, in the absence of other requirements, cannot ensure the quality those job seekers have. As a result, another requirement is established, i.e. applicants must have a background in a particular discipline, making graduates can apply only for a job relevant to their major as required by the company while those whose major is not relevant should apply for a job elsewhere. This may affect the period of time needed by the alumni to get their first job after they graduate.

The data used were obtained from a survey conducted by ITB's Tracer Study Institute undertaken on all ITB alumni Class of 2008. Based on the survey results, four variables were taken, namely *study program* (this research project focused on study programs offered at the Faculty of Mathematics and Natural Sciences and the Faculty of Industrial Engineering) and *waiting period to get the first job after graduation*.

To identify the relationship between those qualitative variables, the statistical method *perceptual mapping* can be used. This method can provide a plot that displays the position on the coordinate axis. Moreover, this method is also used to identify and explain relationships between two variables of the data in the form of a large-dimensional matrix.

Perceptual mapping is commonly performed through several statistical analyses, but Correspondence Analysis was used in this research. According to [1], correspondence analysis is part of the multivariate analysis that investigates the relationship between two or more variables by representing rows and columns simultaneously from a two-way contingency table in a low-dimensional vector space. It is used to reduce dimensions of a variable by searching for the optimal space of the vector space built by the row/ column profile (a 2-dimensional space is usually used) so as to allow the data to be analyzed in depth.

II. RESEARCH METHOD

This research aims to, in general, understand the method of corespondence analysis and to identify the closeness in relationship between the variables understudy thereby exploring as much information as possible from the data and formulating a hypothesis based on the data analysis results.

The data used were obtained from a survey conducted by ITB's Tracer Study Institute on the alumni class of 2008. Based on the data obtained, the variables were divided into a case to be observed in pairs. The case is a pair of variables, namely *Study Program* and *Waiting Time to Get the First Job after Graduation*. Survey results form a matrix/ contingency table with dimensions k x m and then analyzed with corespondence analysis.

III. RESULT AND DISCUSSION

The result of this resarch was divided into two, the fisrt is understand the method of corespondence analysis. And second is aplication example and analysis.

A. Correspondence Analysis

Correspondence Analysis is part of the multivariate analysis that investigates the relationship between two or more variables by representing rows and columns simultaneously from a two-way contingency table in a low/ two-dimensional vector space.

Correspondence Analysis is another form of the principal component analysis, what makes the two different is that the first is used for categorical or qualitative data variables. Like principal component analysis, correspondence analysis is used to reduce a dimensional space into a low-dimensional space and describe as many as possible the structures for the relationship of a data [2]. The correspondence analysis method is a statistical method on the basis of factor analysis. It connects R factor analysis (variable) with Q factor analysis (sample) through transition matrix Z. By the results of R factor analysis, the results of Q factor analysis can be obtained to solve the computational difficulties of large numbers of samples. In the end, the variables and samples are reflected on the same factor plane to reveal the relevance of each variable and each sample as well as that between variables and samples [3].

Basically, correspondence analysis uses the frequency of a contingency table and converts it to a distance, which is then plotted, showing how categories relate one another based on how close or far the distance between one category and another category is in two- or three-dimensional visualization. Nevertheless, this method still has many other functions and the explanation above outlines the main principle on how it works.

For example, a matrix or contingency table with dimensions k x m is obtained, the geometric interpretation of correspondence analysis is to describe distribution of k points across a space with the dimension m (Rm) and distribution of m points across a space with the dimension k (Rk) using a new set of orthogonal linear coordinates which is the principal component, thus the variance of those points is arranged according to their respective value. Therefore, the result of the correspondence analysis is a Cartesian graph which shows the simultaneous projection of each category of the qualitative variables represented by the row and column matrix to the primary axis which explains the largest data variance. The following section describes the process to form a contingency table and the analysis stage of a contingency table, and then presents it as points in a low-dimensional space.

1. Data Construction (Two-Way Contingency Table)

For example, K = {nij | i = 1,2,,k; j = 1,2, ..., m} is a contingency table or cross tabulation that records data obtained from observation by involving two qualitative variables. The qualitative variable X with k categories and qualitative variable Y with m categories.

TABLE I. CONTINGENCY TABLE K

K		**Qualitative Variable Y**				**Total**
		c_1	c_2	…	c_m	
Qualitative Variabel X	r_1	n_{11}	n_{12}	…	n_{1m}	$n_{1.}$
	r_2	n_{21}	n_{22}	…	n_{2m}	$n_{2.}$
	.	.	.	…	.	.
	.	.	.	…	.	.
	.	.	.	…	.	.
	r_k	n_{k1}	n_{k2}	…	n_{km}	$n_{k.}$
Total		$n_{.1}$	$n_{.2}$	…	$n_{.m}$	N

Where:

n_{ij} = the number of objects observed belonging to the category r_k from the qualitative variable X and the category c_m from the qualitative variable Y

$n_{i\bullet}$ = the number of objects observed belonging to the category r_k from the qualitative variable X

$k_{\bullet j}$ = the number of objects observed belonging to the category c_m from the qualitative variable Y

N = total number of observation times

The test used to determine whether there is a relationship between two category variables or not in the form of a contingency table is Pearson's Chi-Square Test. Beased on [4] The statistical test used is described below as Eq. 1:

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{\left(n_{ij} - \widetilde{n_{ij}}\right)^2}{\widetilde{n_{ij}}} \quad (1)$$

where, $\widetilde{n_{ij}}$ = estimated expected value,
degree of freedom ($r_k - 1$) x ($c_m - 1$).

To get visualization of rows and columns from a matrix of the original data to lower dimensions, the matrix $P_{(mxn)}$ should be established first, as the correspondence analysis matrix $P_{(mxn)}$. Matrix P is defined as a relative-frequency matrix obtained by dividing each element n_{ij} by the number of observation times N.

$P_{ij} =$, it r $\frac{n_{ij}}{N}$ efers $\frac{n_{ij}}{N}$ the probability for the object observed belonging to the category r_k from the qualitative variable X and to the category c_m from the qualitative variable Y.

$p_i = \frac{n_i . n_i}{NN}$, it refers the probability for the object observed belonging to the category r_k from the qualitative variable X.

$p_j = \frac{n_{.j} n_{.j}}{NN}$, it refers the probability for the object observed belonging to the category c_m from the qualitative variable Y.

$$P(X = k\ , Y = m) = \frac{P(X,Y)}{p(Y)} = \frac{n_{.ij}}{n_{.j}} = \frac{p_{ij}}{p_j}$$

Thus:

$$\frac{P(X,Y)}{p(Y)} = \frac{n_{.ij}}{n_{.j}} = \frac{p_{ij}}{p_j} \quad (2)$$

is the conditional probability of the object observed belonging to the category r_k from the qualitative variable X if it is found out to belong to the category c_m from the qualitative variable Y.

$$P(\ Y = m, X = k\ ) = \frac{P(X,Y)}{p(Y)} = \frac{n_{ij}}{n_i} = \frac{p_{ij}\, P(X,Y)}{p_{i.}\, p(Y)} = \frac{n_{ij}}{n_{i.}} = \frac{p_{ij}}{p_{i.}} \quad (3)$$

is the conditional probability of the object observed belonging to the category c_m from the qualitative variable Y if it is found out to belong to the category r_k from the qualitative variable X.

Therefore, a table of probability based two-dimensional relative frequency can be established as shown in Table 2.

TABLE II. GENERAL FORM OF TWO-DIMENSIONAL RELATIVE FREQUENCY

P		**Qualitative Variable Y**				**Total**
		c_1	c_2	…	c_m	
Qualitative Variabel X	r_1	p_{11}	p_{12}	…	p_{1m}	$p_{1.}$
	r_2	p_{21}	p_{22}	…	p_{2m}	$p_{2.}$
	.	.	.	…	.	.
	.	.	.	…	.	.
	.	.	.	…	.	.
	r_k	p_{k1}	p_{k2}	…	p_{km}	$p_{k.}$
Total		$p_{.1}$	$p_{.2}$	…	$p_{.m}$	1

With the following matrix P is Eq. 2:

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdot & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ . & \vdots & \ddots & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{km} \end{pmatrix} \begin{pmatrix} p_{11} & p_{12} & \cdot & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ . & \vdots & \ddots & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{km} \end{pmatrix}$$

and

$$P = \frac{1}{N} K_{1\frac{}{N}K} \quad (2)$$

Choose the vector $\overrightarrow{x_i x_i} = \begin{pmatrix} p_{i1} \\ p_{i.} \\ p_{i2} \\ p_{i.} \\ \vdots \\ p_{im} \\ p_{i.} \end{pmatrix} \begin{pmatrix} p_{i1} \\ p_{i.} \\ p_{i2} \\ p_{i.} \\ \vdots \\ p_{im} \\ p_{i.} \end{pmatrix} \in\in R^m R^m$ with i = *1,2…,k* to indicate a set of vectors belonging to the category r_k from the variable X, and vector $\overrightarrow{y_j} = \overrightarrow{y_j} = \begin{pmatrix} p_{1j} \\ p_{.j} \\ p_{2j} \\ p_{.j} \\ \vdots \\ p_{kj} \\ p_{.j} \end{pmatrix} \begin{pmatrix} p_{1j} \\ p_{.j} \\ p_{2j} \\ p_{.j} \\ \vdots \\ p_{kj} \\ p_{.j} \end{pmatrix} \in\in$ $R^k R^k$ with *j=1,2…,m* to indicate a set of vectors belonging to the category c_m from the variable Y.

For example:

a. D_k = diag ($p_{i.}$), *i = 1,2,…,m*

$$D_k = \begin{pmatrix} p_{1.} & 0 & . & 0 \\ 0 & p_{2.} & \cdots & 0 \\ . & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{k.} \end{pmatrix} \begin{pmatrix} p_{1.} & 0 & . & 0 \\ 0 & p_{2.} & \cdots & 0 \\ . & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{k.} \end{pmatrix} \quad D_k =$$

using $p_{i.} p_{i.} = \sum_{j=1}^{m} p_{ij} \sum_{j=1}^{m} p_{ij}$ as the central matrix element for *i= 1,2,…,k*

b. D_m = diag ($p_{.j}$), *j = 1,2,…,m*

$$D_m = \begin{pmatrix} p_{.1} & 0 & . & 0 \\ 0 & p_{.2} & \cdots & 0 \\ . & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{.m} \end{pmatrix} \begin{pmatrix} p_{.1} & 0 & . & 0 \\ 0 & p_{.2} & \cdots & 0 \\ . & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{.m} \end{pmatrix}$$

using $p_{.j} p_{.j} = \sum_{i=1}^{k} p_{ij} \sum_{i=1}^{k} p_{ij}$ as the central matrix element for *j= 1,2,…,m*

Thus, the i^{th} row of $U = P^t D_k^{-1}$ indicates profiles of the row in the category r_k of the variable X while the j^{th} row of $V = P D_m^{-1}$ indicates profiles of the column in the category c_m of the variable Y.

2. Singular Value Decompotition

As mentioned previously, the geometric interpretation of correspondence analysis is to describe the distribution of k points across a space with the dimension m (R^m) and distribution m points across a space with the dimension k (R^k) using a new set of orthogonal linear coordinates which are principal components. The principal components are presented as the proportion of the total variance (the sum of all eigenvalues).

To reduce the dimension based on data variance, which is by using the largest eigenvalue/ inertia value to maintain optimal information, it is necessary to employ the method known as *Singular Value Decomposition* (SVD). This method is a process of factoring a matrix into more than one matrix, i.e. multiplication between the diagonal matrix containing singular values and two orthogonal matrices containing corresponding singular vectors.

D = max (k, m) - 1 shows the maximum number of dimensions required to describe the relationship between rows and columns graphically. For example, a contingency table has dimensions of 4 x 6, then D = max (4.6) - 1 = 5, meaning that there are a maximum of 5 dimensions that can be formed to describe the relationship between two qualitative variables [5].

Definition 1:
For example, matrix *A* has dimensions *k* x *m*. The singular value decomposition of the real matrix *A* is the factorization of:

$$A = U\boldsymbol{\Sigma}\, V^t$$

where $U_{kxk} U_{kxk}$ and $V_{mxm} V_{mxm}$ are orthogonal matrices and Σ is a diagonal matrix of *k*x*m* with non-negative reals known as singular values. In other words, $\Sigma = \text{diag}\ (\sigma_1\sigma_1, \sigma_2\sigma_2, \dots, \sigma_k\sigma_k)$ is in descending order and as a result $\sigma_1\sigma_1 \geq \sigma_2\sigma_2 \geq \cdots \geq \sigma_k\sigma_k$.
If $U = (u_1u_1, u_2u_2, \dots, u_nu_n)$ and $V = (v_1v_1, v_2v_2, \dots, v_nv_n)$, thus required Eq. 3:

$$A = \sum_{i=1}^{n} \sigma_i u_i v_i$$

The definition also suggest that the matrix $A_{(k\ x\ m)}$ can be expressed as the decomposition of matrixes, namely matrixes *U, Σ,* and *V*. The matrix *Σ* is a diagonal matrix with diagonal elements in the form of singular values of the matrix A while matrices U and V are matrices whose columns consist of the right and left singular vectors of the matrix A, for the corresponding singular values.

To determine SVD includes the steps to determine eigenvalues and eigenvectors of the matrix AA^t or A^tA. The eigenvector of A^tA forms the column V while the eigenvector of AA^t forms the column U. The singular value in Σ is the square root of the eigenvalues of the matrix AA^t or A^tA. Singular values are diagonal elements of Σ and arranged in descending order.

Definition 2:
A matrix with rank *r* is given. The positive eigenvalue of $(A^tA)^{1/2}$ is called the singular value of A. In other words, if *σ* is the singular value of A, it means that *σ* is the positive eigenvalue of $(A^tA)^{1/2}$, σ^2 is the eigenvalue of A^tA.

Based on the definition above, the relationship between the eigenvalue and the singular value can be determined. In other words, for the matrix A with rank *r* and eigenvalues of the matrix A^tA, $\lambda_1\lambda_1 \geq \lambda_2\lambda_2 \geq \cdots \geq \lambda_r\lambda_r \geq \lambda_{r+1}\lambda_{r+1} = \cdots = \lambda_n\lambda_n = 0$, therefore $\sigma_i\, \sigma_i = \sqrt{\lambda_1}\sqrt{\lambda_1}$ with *i = 1, 2, …, r, r*+1, …, *n* is called the singular value of the matrix A.

In Correspondence Analysis, the matrix to which the method *SVD* will be performed in the column profile analysis is the data matrix *T1* (with dimensions *k* x *m*), which is a non-centralized data matrix that later will be factorized into Eq. 4:

$$T_1 = VD_mV^t \qquad (4)$$

with *rank* (T_1) = min ((*k* – 1) (*m* – 1)), V as an orthogonal matrix with dimensions *k* x *k*, as an orthogonal matrix with dimensions *m* x *m*, and D_m as a diagonal matrix with dimensions *k* x *m*.

As for row profile analysis, the matrix to which the method *SVD* will be performed is the data matrix S_1 (with dimensions *k* x *m*), which is a non-centralized data matrix that later will be factorized into Eq. 5:

$$S_1 = UD_kU^t \qquad (5)$$

with *rank* (S_1) = min ((*k* – 1) (*m* – 1)), *U* as an orthogonal matrix with dimensions *k* x *k*, U^t as an orthogonal matrix with dimensions *m* x *m*, and D_k as a diagonal matrix with dimensions *k* x *m*.

In Correspondence Analysis, it is unnecessary to centralize data matrices *U* and *V* as the same analysis results will be generated if centralized data matrices are used.

3. Determination of the Distance between Profiles

The distance between two categories was not calculated using a formula to measure the Euclidean distance, but using the Chi-Square distance. Because using the Euclidean distance, the combination of two categories of the column profile will result in a change in the distance between two categories of the row profile, and vice versa. Conversely, using the chi-square distance, the combination of two categories of the column profile will not result in a change in the distance between two categories of the row profile, and vice versa [6].

Consistent with the foregoing,[7] state that regardless of which dimension reduction techniques used by analysts, it is important that every single similarity, or difference, between profiles is reflected in a low-dimensional plot. Such a rule is referred to as Equivalence in Distribution. If two profile rows, or columns, have identical combined profile distribution, the distance between them does not change.

In column profile analysis, determination of the distance between 2 categories $\overrightarrow{y_jy_j}$ and $\overrightarrow{y_jy_j}$ is based on the chi-square distance M = D_k^{-1}. For more details, refer to Theorem 2 in the Appendix. As for row profile analysis, determination of the distance between 2 categories $\overrightarrow{x_ix_i}$ and $\overrightarrow{x_ix_i}$ is based on the chi-square distance N = D_m^{-1}.

4. Relationship between the Column Analysis at R^k and the Row Analysis at R^m

The relationship between the column analysis and the row analysis will be described below.

As mentioned earliear, processes in the row profile and column profile analyses will generate the same eigenvalue. For more details, refer to Theorem 11 in the Appendix.

For example $1 = \lambda_1 = \cdots = \lambda_{q-1} > \lambda_q \geq \cdots \geq \lambda_m \geq 0$
are eigenvalues of $T_1\, D_k^{-1}$. Thus, vectors for categories *X* and *Y* of the main plane P_1 at R^k should be presented as follows:

1) $\overrightarrow{y_j y_j}$ presented by the vector $\overrightarrow{\varphi_j \varphi_j} = \begin{pmatrix} a_j^q \\ a_j^{q+1} \end{pmatrix}\begin{pmatrix} a_j^q \\ a_j^{q+1} \end{pmatrix}$, with $j=1,2,\ldots,m$
2) $\overrightarrow{x_i x_i}$ presented by the vector $\overrightarrow{y_i y_i} = \begin{pmatrix} \frac{1}{\sqrt{\lambda_q}} \overrightarrow{b_i^q} \\ \frac{1}{\sqrt{\lambda_{q+1}}} \overrightarrow{b_i^{q+1}} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{\lambda_q}} \overrightarrow{b_i^q} \\ \frac{1}{\sqrt{\lambda_{q+1}}} \overrightarrow{b_i^{q+1}} \end{pmatrix}$, with $i=1,2,\ldots,k$

Therefore, in P_1 the following closeness can be observed:

a. Between categories of the variable *X* by observing { $\overrightarrow{\varphi_j \varphi_j}$ | j = 1,…,m }
b. Between categories of the variable *Y* by observing { $\overrightarrow{y_i y_i}$ | i = 1,…,k }
c. Between a category of the variable *X* and categories of the variable *Y* by observing $\overrightarrow{\varphi_j \varphi_j}$ and $\overrightarrow{y_i y_i}$, for $i = 1,\ldots,k$ and $j = 1,\ldots,m$.

In simultaneous presentation in P_1, it is revealed that. $p_{ij}p_{ij} = p_{.j}p_{.j}$ It means that for h ≠ I, the value of $p_{hj}p_{hj} = 0$ thus $a_j^q a_j^q = \frac{1}{\sqrt{\lambda_q}}\overrightarrow{b_i^q}\frac{1}{\sqrt{\lambda_q}}\overrightarrow{b_i^q}$. As a result, $\overrightarrow{\varphi_j \varphi_j} = \overrightarrow{y_i y_i}$ or in other words, the category $c_m c_m$ of the variable *Y* and the category $r_k r_k$ of the variable *X* are in close proximity to each other.

5. Contribution of the Row Profile and the Column Profile

Among the methods to assess results of correspondence analysis is by looking at the value of inertia contribution generated by the primary axes. If the first two axes generate an inertia value that is fairly high, this means that both first two primary axes can represent information and ignore the other primary axes by not causing significant information loss. However, if the majority of the percentage of total inertia lies in another primary axes, it means that there are some points (categories) that cannot be displayed well by both first primary axes.

The inertia value of an axis can be calculated by having the singular value squared, the value equals the square of the distance between the point and the axis center weighted by the mass of each point. The square of the distance between the point and the axis center weighted can be defined as the percentage of the root of characteristics, which is knows as absolute contribution or contribution of points to the root of characteristics or to the primary axes.

1) Absolute Contribution

Absolute contribution indicates the proportion of the variance which each category can explain in connection with the primary axes.

The absolute contribution of the i^{th} row profile is defined as Eq. 6 follows:

$$RAC = \frac{p_{i\cdot}(a_i^\alpha)^2}{\lambda_\alpha}\frac{p_{i\cdot}(a_i^\alpha)^2}{\lambda_\alpha} \text{ for } i=1,\ldots,k,\ \alpha\alpha=1,\ldots,m \quad (6)$$

2) The absolute contribution of the j^{th} column profile is defined as Eq. 7 follows:

$$CAC = \frac{p_{.j}\cdot(b_j^\alpha)^2}{\lambda_\alpha}\frac{p_{.j}\cdot(b_j^\alpha)^2}{\lambda_\alpha} \text{ for } i=1,\ldots,m,\ \alpha\alpha=1,\ldots,m \quad (7)$$

It can be concluded that points with a greater mass value or situated more distant from the center of the axes may generate greater inertia contribution. As the limit to indicate whether or not the contribution of each category is deemed good, the criterion below is commonly used:

$$RAC(CAC) \geq \frac{100}{number\ of\ rows\ (coloumns)}$$

If $RAC(CAC) \leq \frac{100}{number\ of\ rows\ (coloumns)}$, it means that the category does not have sufficient inertia contribution for each dimension/ primary axis.

3) Relative Contribution

Relative contribution indicates representation quality of each row/ column category of the $\alpha\alpha^{th}$ axis. The relative contribution of the i^{th} row profile is defined as follows:

$$RRC_i = \frac{(a_i^\alpha)^2}{\sum_{\alpha=1}(a_i^\alpha)^2}, \text{ for } i=1,\ldots,k,\ \alpha=1,\ldots,m.$$

The absolute contribution of the j^{th} column profile is defined as follows:

$$CRC_j = \frac{(b_j^\alpha)^2}{\sum_{\alpha=1}(b_j^\alpha)^2}, \text{ for } i=1,\ldots,m,\ \alpha=1,\ldots,m$$

The higher the value of relation contribution of the row/ column profile is, the better the representation quality of the profile of the α^{th} axis is. Relative contribution can be defined as the square of the cosine of an angle formed by each vector of the row/ column profile with the projection vector of the α^{th} axis. A higher square of the correlation value indicates that it has a better ability in explaining the inertia value, and vice versa, the smalled the square of the correlation value is, the smaller the inertia value that can be explained by the primary axis [8].

B. Aplications Example and Analysis

The case study in this research was perfomed by using data obtained from the survey undertaken by ITB's Tracer Study Institute on students Class of 2008 with a total of 2,612 respondents filling in questioners distributed. Based on results of the survey, two variables were taken by observing alumni from the Faculty of Mathematics and Natural Sciences and the Faculty of Industrial Engineering only consisting of as many as 485 persons, namely the variables *Study Program* and *Waiting Period to Get the First Job after Graduation*.

Correspondence Analysis between the Variables *Study Program* and *Waiting Period to Get the First Job after Graduation*

Each variable was broken down into the following categories:

a) The Variable *Study Program* (X)

The study programs under study were those of the Faculty of Mathematics and Natural Sciences and the Faculty of Industrial Engineering, namely:

The following are study programs at the Faculty of Mathematics and Natural Ssciences:

(1) Astronomy
(2) Physics
(3) Chemistry
(4) Mathematics

The following are study programs at the Faculty of Industrial Engineering:

(5) Chemical Engineering
(6) Industrial Engineering
(7) Engineering Physics

b) The Variable *Waiting Period to Get the First Job after Graduation* (Y)

Thus variable was broken down into 4 categories, namely:

(1) : Less than 3 months
(2) : 3-6 months
(3) : 7-12 months
(4) : More than 12 months

Stages of Correspondence Analysis:

1. Drawing a Contingency Table

Cross tabulation (contingency table) between the variables *Study Program* and *Waiting Pperiod to Get the First Job after Graduation* of ITB alumni class of 2008 from the Faculty of Mathematics and Natural Sciences and the Faculty of Industrial Engineering is presented in Table 3.

It can be seen in Table 3 that the total number of alumni the Faculty of Mathematics and Natural Sciences and the Faculty of Industrial Engineering working after their graduation is as many as 362 persons, the total frequency for all rows of the variable *Study Program* for the study program *Industrial Engineering* is 94 while the total frequency for all rows of the variable *Waiting Period to Get the First Job after Graduation* for the category 2 ((3; 6] bulan) is 162.

TABLE III. CONTINGENCY TABLE (CORRESPONDENCE) BETWEEN VARIABLES STUDY PROGRAM AND WAITING PERIOD TO GET THE FIRST JOB AFTER GRADUATION.

Study Program	Waiting period to get the first job after graduation				
	(0;2] months	(3;6] months	(7;12] months	>12 months	Active margin
Astronomy	1	4	0	1	6
Physics	15	15	7	0	37
Chemistry	9	26	4	9	48
Mathematics	21	27	9	5	62
Chemical Engineering	20	33	8	2	63
Engineering Physics	27	17	6	2	52
Industrial engineering	43	40	9	3	94
Active margin	135	162	43	22	362

2. Chi-Square Test between the Variables *Study Program* and *Waiting Period to Get the First Job after Graduation*

As mentioned previously, the correspondence analysis was undertaken using a statistical test, namely chi-square test to explain the total variances and to examine the level of significance of the relationship between the two variables. Based on Table 4, it is revealed that the chi-square value generated is equal to 37.280 while the p-value is equal to 0.005, meaning that there is a relationship between the variables *Study Program* and *Waiting Period to Get the First Job after Graduation*. Thus, in order to further determine patterns for the relationship between both variables, correspondence analysis may be performed.

3. Correspondence Analysis Results

In correspondence analysis, the output of SPSS 19 displays only dimensions that can be interpreted rather than displays all the dimensions that explain a particular variable. Dimension 1 will always explain the greatest variance, then Dimension 2, and so on.

TABLE IV. TABLE OF INERTIA AND PROPORTION OF VARIANCES OF THE VARIABLES *STUDY PROGRAM* AND *WAITING PERIOD TO GET THE FIRST JOB AFTER GRADUATION*.

Dimension	Singular value	inertia	Chi square	Sig.	Proportion of inertia	
					Accounted for	cumulative
1	.286	.082			.794	.794
2	.118	.014			.135	.929
3	.085	.007			.071	1.000
Total		.013	37.280	.005[a]	1.000	1.000

In Table 4, there are three dimensions formed but only two will be interpreted. The column '*inertia*' shows the total variance that can be explained by each dimension. It can be seen in Table 4 that the total inertia (the total variance that can be explained) is equal to 10.3%. This indicates that based on results of the correspondence analysis, the variable *Study Program* can explain the variable *Waiting Period to Get the First Job after Graduation*, which is as many as 10.3% and vice versa.

The column *singular value* shows the root of the eigenvalue. The singular values are equal to 0.286, 0.118, and 0.085 for the first dimension (the highest), the second dimension (the second highest), and the third dimension, respectively. The values of each dimension was put in ascending order based on the percentage of the variance described in the analysis results in the column *Proportion of Inertia*. Two dimensions (the primary axis) obtained generate representation quality by 92.9% (79.4% + 13.5%), which implies that only 2 dimensions are used as those dimensions have already been able to explain data variance very well, which is by 92.9%.

As mentioned earlier, one of the methods to assess results of correspondence analysis is to look at the value of inertia contribution given by the primary axis. In addition, the

representation quality of each row/ column category on the a^{th} axis through the relative contribution (Contribution of Point to Inertia of Point) can also be determined. The absolute contribution and relative contribution generated from the analysis of the row profile and the analysis of the column profile undertaken separately are described below.

Row Profile Analysis Results (the Variable *Study Program*)

Based on Table 3, there are 4 columns of categories for the Waiting Period to Get the First Job after Graduation and as a result, the Row Profile Analysis should be undertaken at R^4 which will provide which will provide information about the way the variable *Study Program* will be made in the row analysis plot. In the output obtained using SPSS 19, there are coordinates to be used to put points of each category in the byplot for the row analysis and the byplot for the joint analysis of the rows and columns.

The column *Score in Dimension* shows coordinates of each point (study program) in its dimensions (1 and 2) to be put in the plot (biplot) of the row analysis. These coordinates are eigenvectors that correspond to eigenvalues and form the principal component that represents rows of a data matrix.

Dimensions (the primary axes) can be interpreted by looking at contribution of each category (*Study Program*) to each dimension. From Table 5, it can be seen that there are 7 categories (study programs), meaning that contribution higher that the absolute contribution criterion, which is 100/7 = 14.29 %, will suggest that the category provides major contribution in explaining variance in data for each dimension.

TABLE V. COORDINATES AND CONTRIBUTION OF THE ROW PROFILE (THE VARIABLE *STUDY PROGRAM*). IT IS REVEALED THAT THE VARIABLE *STUDY PROGRAM* FOR THE STUDY PROGRAM *PHYSICS* HAS TWO COORDINATES, THEY ARE 0.509 IN DIMENSION 1 AND 0.440 IN DIMENSION 2.

Study program	Score in dimension		Contribution				
	1	2	Of point to inertia of dimension		Of dimension to inertia of point		
			1	2	1	2	Total
Astronomy	-1.1250	-.037	.091	.000	.852	.000	.853
Physics	.509	.044	.092	.168	.677	.209	.886
Chemistry	-1.157	-.159	.062	.029	.989	.008	.997
Mathematics	-.138	.054	.011	.004	.344	.022	.366
Chemical Engineering	-.022	.510	.000	.384	.004	.859	.863
Engineering Physics	.468	-.537	.110	.352	.639	.347	.986
Industrial engineering	287	-.170	.075	.063	.696	.100	.796
Total			1.000	1.000			

Based on Table 5, the point *Chemistry* (62%) meets this criterion. This means that Chemistry is the only study program that provide significant contribution to explain data variance in Dimension 1 (the first primary axis) while other study programs give small contribution only.

For Dimension 2, points for Physics (16.8%), Chemical Engineering (38.4%), and Engineering Physics (35.2%) provide significant contribution to explain Dimension 2 (the second primary axis). This means that the study programs *Physics*, *Chemical Engineering*, and *Engineering Physics* provide significant contribution to explain data variance in Dimension 2 (the second major axis) while other study programs give small contribution only.

The representation quality of each category (study program) can be examined from the value of relative contribution (Contribution of Point to Inertia of Point) which indicates representation quality of each category in the primary axis. The closer the contribution value is to the value 1 is, the better the representation quality of the category on the primary axis is.

From Table 5, it can be seen that for Dimension 1, the values of relative contribution generated for study programs *Astronomy* (0.852) and *Chemistry* (0.989) are close to 1. It means that the representation quality of the study programs *Astronomy* and *Chemistry* in Dimension 1 is good while the representation quality of the other study programs is less good in Dimension 1. As for Dimension 2, the relative value generated by the study program *Chemical Engineering* is equal to 0.859, meaning that *Chemical Engineering* is the only study program with good representation quality in Dimension 2 while the other study programs have less good representation quality in Dimension 2. After obtaining coordinates for each point, the next step is to place those coordinates in the plot as shown in Figure 1.

Column Profile Analysis Results (the Variable *Waiting Period to Get the First Job after Graduation*)

Similar to the row analysis, in Table 3 there are 7 rows of study program categories, therefore the Column Profile Analysis should be performed at R^7 which will provide information about the way the variable *Waiting Period to Get the First Job after Graduation* will be made in the plot. These coordinates are eigenvectors that correspond to eigenvalues and form the principal component that represents the column of a data matrix.

Dimensions (the primary axes) are interpreted by looking at contribution of each category (*Waiting Period to Get the First Job after Graduation*) to each dimension. From Table 6, it can be seen that there are 4 categories, meaning that contribution higher that the absolute contribution criterion, which is 100/4 = 25 %, will suggest that the category provides major contribution in explaining variance in data for each dimension.

Based on Table 6, *Waiting Period to Get the First Job after Graduation* in Categories 1 (29%) and 4 (59.2%) meet the ctiterion. It means that only *Waiting Period to Get the First Job after Graduation* in Categories 1 and 4 that provide significant contribution in explaining data variance in Dimension 1 (the first primary axis) while *Waiting Period to*

Get the First Job after Graduation in Categories 2 and 3 only give small contribution.

TABLE VI. COORDINATES AND CONTRIBUTION OF THE COLUMN PROFILE (THE VARIABLE *WAITING PERIOD TO GET THE FIRST JOB AFTER GRADUATION*). IT IS REVEALED THAT COORDINATES FOR THE VARIABLE *WAITING PERIOD TO GET THE FIRST JOB AFTER GRADUATION* FOR CATEGORY 2 ((3; 6] MONTHS)) ARE -0.237 FOR DIMENSION 1 AND 0.245 FOR DIMENSION 2.

Waiting period to get the first job after graduation	Score in dimension		Contribution				
	1	2	Of point to inertia of dimension		Of dimension to inertia of point		
			1	2	1	2	Total
(1) (0;2] months	.472	-.325	.290	.334	.836	.163	.999
(2) (3;6] months	.237	.245	.088	.227	.596	.262	.857
(3) (7;12] months	.265	.446	.029	.200	.242	.281	.522
(4) >12 months	-1.670	.681	.592	.239	.922	.063	.985
Active total			1.000	1.000			

Then for Dimension 2, *Waiting Period to Get the First Job after Graduation* in Category 1 (33.4%) contributes to the establishment of Dimension 2 (the second primary axis). It means that only *Waiting Period to Get the First Job after Graduation* in Category 1 provides significant contribution in explaining data variance in Dimension 2 (the second primary axis) while *Waiting Period to Get the First Job after Graduation* in Categories 2, 3 and 4 only give small contribution.

Relative contribution (*Contribution of Point to Inertia of Point*) shows representation quality of each category (program studi) in the primary axes. The closer the value of relative contribution to 1 is, the better the representation quality of a category in the primary axes is. Based on Table III.4, it is revealed that for Dimension 1, the value of relative contribution for *Waiting Period to Get the First Job after Graduation* in Categories 1 (0.836) and 4 (0.922) generated is close to 1. It means that the representation quality of *Waiting Period to Get the First Job after Graduation* in Categories 1 and 4 in Dimension 1 is good while *Waiting Period to Get the First Job after Graduation* in Categories 2 and 3 has less good representation quality in Dimension 1.

Afterwards, for Dimension 2, the relative value of *Waiting Period to Get the First Job after Graduation* for all categories is getting increasingly far from 1, meaning that the representation quality of the waiting period to get a job for all categories in Dimension 2 are not good enough.

After coordinates for each point have been discovered, the next step is to put those coordinates in the plot as illustrated in Figure 1

Results of the Simultaneous Analysis of the Row Profile and the Column Profile

Based on Figure 1 it can be seen that the coordinates falling into Category 1 ((0;2] months) of the variable *Waiting Period to Get the First Job* is in close proximity to the coordinates of Industrial Engineering and Engineering Physics of the variable *Study Program*.

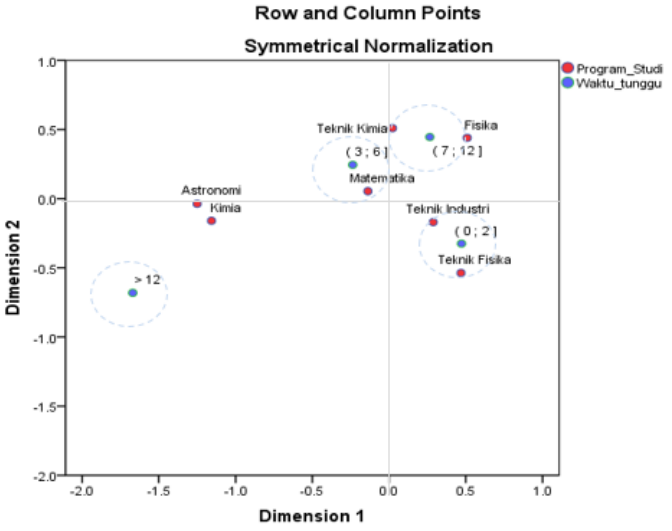

Fig. 1. Biplot illustrating correspondence analysis for the variables *Study Program* and *Waiting Period to Get the First Job after Graduation*.

Figure 2 shows the output of correspondence analysis in the form of a biplot of the joint analysis of rows and columns undertaken simultaneously. As mentioned previously, *score in dimension* of each row analysis and column analysis was used to form the biplot. Based on Table 4, Dimension 1 can explain variance in data by 79.4%, while Dimension 2 can explain variance in data by 13.5%.

The adjacent categories in Figure 1 show that those categories have similarities. The circles in Figure 1 are the criteria used to make groups based on proximity between points. Using the variable *Waiting Period to Get the First Job after Graduation* as a central point with a diameter of 0.5 cm, the points of the variable *Study Program* adjacent to the points of the variable *Waiting Period to Get the First Job after Graduation* are considered as one group.

If the proximity of each of the points in Figure 1 and Table 3 is observed, the following conclusions can be drawn:

1. The points for the study programs *Industrial Engineering* and *Engineering Physics* are in fairly close proximity to Category 1 ((0; 2] months) for the variable *Waiting Period to Get the First Job after Graduation*. This indicates that ITB alumni Class of 2008 majoring in Industrial Engineering and Engineering Physics generally get their first job the fastest among other study programs, i.e. in less than two months after their graduation.
2. The point for the study program *Mathematics* is in fairly close proximity to Category 2 ((3; 6] months) for the variable *Waiting Period to Get the First Job after Graduation*. This indicates that ITB alumni Class of 2008

majoring in Mathematics generally need 3 to 6 months to get their first job the fastest among other study programs.

Based on description in points 1 and 2, it is evident that those study programs do have good job prospects and are needed badly by companies to improve their quality and products, making it relatively fast for them to get their first job after graduation.

3. The distance between the point of the study program *Chemical Engineering* and the point of Category 3 ((7, 12] months) for the variable *Waiting Period to Get the First Job after Graduation* is close to each other. But, the distance between the point of the study program *Chemical Engineering* is closer to the point of Category 3 ((7, 12] months) for the variable *Waiting Period to Get the First Job after Graduation* and so is the study program *Physics*. This indicates that ITB alumni Class of 2008 majoring in Chemical Engineering and Physics generally need 7 to 12 months to get their first job after their graduation.
4. The study programs *Astronomy* and *Chemistry* are in close proximity to Category 4 (> 12 months) for the variable *Waiting Period to Get the First Job after Graduation*. This indicates that ITB alumni Class of 2008 majoring in Astronomy and Chemistry generally require the longest waiting period to get a job among other study programs, i.e. within more than 12 months after their graduation.

Whereas from points 3 and 4, alumni from study programs *Chemical Engineering*, *Astronomy*, and *Chemistry* take a long time to get their first job after graduation. Such a condition may result from several things, one of which is lack of employment opportunities available at the time they graduated or it may also happen because the alumni wait for a better job (a hig-paying one) and various other reasons that cause alumni from the study programs to take a long time to get a job after their graduation.

Findings of this research can be used to gain information by prospective university students to choose a study program that offer them a great work opportunity and does not take long to get a job.

IV. CONCLUSIONS

To process the data obtained from measurement of two qualitative variables in the form of categories, a tool called Correspondence Analysis can be used. Correspondence analysis aims to observe a data matrix that is universal, complex, and multidimensional in nature, then reduce the matrix to a low-dimensional matrix by dividing it into smaller parts to make it easier to analyze and interpret each part, then re-combine them thus allowing interpretation of the data to be observed.

Results of the survey undertaken by ITB's Tracer Study Institute in 2008, for variables *Study Program* and *Waiting Period to Get the First Job after Graduation* suggest that:

1. Study programs with a period of waiting to get the first job after graduation by less than six months are Industrial Engineering, Engineering Physics, and Mathematics.
2. Study programs with a period of waiting to get the first job after graduation ranging from seven to twelve months are Chemical Engineering and Physics.
3. Lastly, study programs with the longest period of waiting to get the first job after graduation, which is by more than 12 months, are Astronomy and Chemistry.

REFERENCES

[1] M. Greenacre, *Theory and Application Of Correspondence Analysis*. Academic Press London, 1984.
[2] E.J. Beh and R. Lombardo, *Correspondence analysis. Theory, practice and new strategies*, vol. 53, no. 9. Australia: John Wiley & Sons, Ltd, 2014.
[3] X.D. Xiang Shiyao, Zhao Mingdong, Chen Yijin, "Utilization Strategy Research Of Goafs Based On Correspondence Analysis Shiyao Xiang," no. Aeecs, pp. 52–56, 2016.
[4] J. Wichern;, *Applied Multivariate Statistical Analysis*, Fifth Edit. Prentice Hall,Inc, 2002.
[5] E.J. Beh, "Simple correspondence analysis: a bibliographic review," *Int. Stat. Rev.*, vol. 72, no. 2, pp. 257–284, 2004.
[6] A. Sukmana, "Analisis Korespondensi: Identifikasi Profil Latar Belakang Pendidikan Orang Tua Terhadap Profil Prestasi Akademik Mahasiswa ITB," Institut Teknologi Bandung, 1992.
[7] Aryfianto, "Analisis Korespondensi," Institut Teknologi Bandung, 2005.
[8] A. Agresti, *An introduction to categorical Data Analysis*, Second., vol. 22, no. 1. Gainesvile, florida: John Wiley & Sons, Ltd, 1996.