Volume 12
Number 1
2020

# Internetworking Indonesia Journal

**The International Journal of ICT and Internet Development**



ISSN: 1942-9703

www.InternetworkingIndonesia.org

# Internetworking Indonesia Journal

The International Journal of ICT and Internet Development
ISSN: 1942-9703
www.InternetworkingIndonesia.org


The *Internetworking Indonesia Journal* (IIJ) is a peer reviewed international journal devoted to the timely study of Information and Communication Technology (ICT) and Internet development. The journal seeks high-quality research papers on the challenges and opportunities presented by information technology and the Internet, as well as their impact to society. Issues in the IIJ covers the broad areas of computing, communications, engineering (software & hardware) and the broader area of internet technologies.

The journal provides open access to all of its content on the principle that making research freely available to the public supports a greater global exchange of knowledge. This follows the philosophy of the Open Journal Systems (see the Public Knowledge Project at pkp.sfu.ca). The IIJ issues are published electronically (PDF) and there are no subscription fees. Such access is associated with increased readership and increased citation of an author's work.


Journal mailing address: Internetworking Indonesia Journal, PO Box 397110 MIT Station, Cambridge, MA 02139, USA.


## Computing, Communications, Engineering & Internet Technologies

The Internetworking Indonesia Journal covers the broad technical areas of computing, communications, engineering and Internet Technologies.

Possible topics for papers include, but are not limited to the following:

- Information technology and information systems
- Communications technology
- Software and hardware engineering
- Algorithms and computation
- Applications and services
- Broadband and telecommunications technologies
- Mobile and wireless networks
- Internet infrastructure systems, protocols, standards and technologies
- Multimedia and content development
- Education and distant learning
- Open source software development and deployment
- Cloud Computing, SaaS, and Grid Computing

# Internetworking Indonesia Journal

The International Journal of ICT and Internet Development
ISSN: 1942-9703
www.InternetworkingIndonesia.org


The *Internetworking Indonesia Journal* (IIJ) is a peer reviewed international journal devoted to the timely study of Information and Communication Technology (ICT) and Internet development. The journal seeks high-quality research papers on the challenges and opportunities presented by information technology and the Internet, as well as their impact to society. Issues in the IIJ covers the broad areas of computing, communications, engineering (software & hardware) and the broader area of internet technologies.

The journal provides open access to all of its content on the principle that making research freely available to the public supports a greater global exchange of knowledge. This follows the philosophy of the Open Journal Systems (see the Public Knowledge Project at pkp.sfu.ca). The IIJ issues are published electronically (PDF) and there are no subscription fees. Such access is associated with increased readership and increased citation of an author's work.


Journal mailing address: Internetworking Indonesia Journal, PO Box 397110 MIT Station, Cambridge, MA 02139, USA.


## Publication Policies

**Open Access Publication Policy**: The Internetworking Indonesia Journal (IIJ) provides open access to all of its content on the principle that making research freely available to the public supports a greater global exchange of knowledge. This follows the philosophy of the Open Journal Systems (see the Public Knowledge Project at pkp.sfu.ca). The IIJ issues are published electronically (PDF) and there are no subscription fees. Such access is associated with increased readership and increased citation of an author's work.

**No publication fees**: There is no publication fees to authors for publishing in the IIJ. This is consistent with the open access philosophy adopted by the IIJ.



## Manuscript Submission Guidelines

**Language**: The primary language of the Internetworking Indonesia Journal is English due to its international readership. Note that the IIJ does not accept requests for the translation of submitted manuscripts.

**Format**: Papers should be submitted using the IIJ Template document for MS-Word. The IIJ format is derived from the popular IEEE 2-column text format. The IIJ template can be found on the IIJ website.

**Length of manuscript**: Papers submitted to the IIJ should not exceed 7 pages of the IIJ 2-column format.

**Originality**: Papers submitted to the IIJ must not have been previously published or committed to another publisher under a copyright transfer agreement, and must not be under consideration by another journal. Papers previously published at conferences can be submitted to the IIJ, but must be revised so that it has significant differences from the conference version.

**Ethics Statement**: Authors wishing to submit papers to the IIJ are requested to read and agree to the IIJ *Publication Ethics Statement*, which can be found on the IIJ website.

# Correspondence Analysis on Tracer Study Data Using Confidence Circle (Study Case: Alumni of Institut Teknologi Bandung, Batch 2008 and 2009)

U. S. Pasaribu, N. F. F. Ilmi, U. Mukhaiyar and H. Husniah

*Abstract*— **As the second-best university in Indonesia, Institut Teknologi Bandung has been carrying out tracer studies since 2010. It is very important because the institution can get feedback from alumni to improve learning method or can manage programs to upgrade student quality. Correspondence analysis is one method to obtain interesting patterns from a set of data through some graphical displays. This analysis works by projecting the row and column profiles into the subspace simultaneously, in order to produce two or three Euclid subspaces. The dimension reduction technique that applied is the singular value decomposition. In this paper, the correspondence analysis is used to find out the predicate (student's academic achievement) trend pattern from two different categories in the tracer study. The final step in correspondence analysis is to complete the map with a confidence circle. This circle can measure the proximity of coordinates between several categories. It can also identify categories that have significant influences. As an illustration, two categorical variables of tracer study used, i.e., grade point cumulative and waiting time to get first job data for a different batch of alumni. It is obtained that generally, all faculties have similar behavior. But for some faculties, there is no significant tendency between both variables. It can be seen from the correspondence map with 10% confidence region still cannot give significant influence. It is concluded that, in order to reach the quality of alumni, the university should give more soft skills (some soft skills out of academic skills) or career guidance for students, especially for students whose academic achievement are lower than Very Satisfied.**

***Index Terms*—Multivariate analysis, correspondence analysis, singular value decomposition, tracer study, circle confidence**.

Manuscript received November 1, 2019. This research was supported in part by the Programme of Research and Community Service (P3MI) for Institut Technology Bandung in 2019. One of the research results is a paper entitled Correspondence Analysis on Institut Teknologi Bandung Tracer Study Data Using Confidence Circle (Study Case: Alumni of Batch 2008 and 2009).

U.S. Pasaribu is an associate Professor and author for correspondence for this paper. She is now with Statistics Research Division in the Mathematical Programme Study, Faculty of Mathematics and Natural Science, Institut Teknologi Bandung (e-mail: udjianna@math.itb.ac.id).

N.F.F. Ilmi is a graduate magister students from Mathematical Programme Study, Institut Teknologi Bandung.

U. Mukhaiyar is with Statistics Research Division in the Mathematical Programme Study, Faculty of Mathematics and Natural Science, Institut Teknologi Bandung (e-mail: utriweni@math.itb.ac.id).

H. Husniah is with Industrial Engineering, Langlangbuana University, Indonesia. (e-mail: hennie.husniah@gmail.com).

## I. Introduction

Tracer study is an approach that enables higher education institutions (HEIs) to obtain information about possible deficiencies in educational and learning process and form the basis for planning activities for the improvement in the future [1]. Also, referred to as graduate or alumni survey, tracer study is a graduate survey that attempts to draw out graduates' activities or previous students of an educational institution. It is aimed to determine the status of graduates within a certain period of time after completing their studies. Furthermore, the study was conducted by institutions of education to obtain feedback from alumni. It needs to be executed so that the institution can improve the quality of the output and improve the education system [2]. In Indonesia, most of best universities have done this tracer study since 4-6 years ago, such as Institut Teknologi Bandung, Universitas Indonesia, Universitas Gadjah Mada, Institut Teknologi Surabaya, Universitas Padjadjaran, etc.

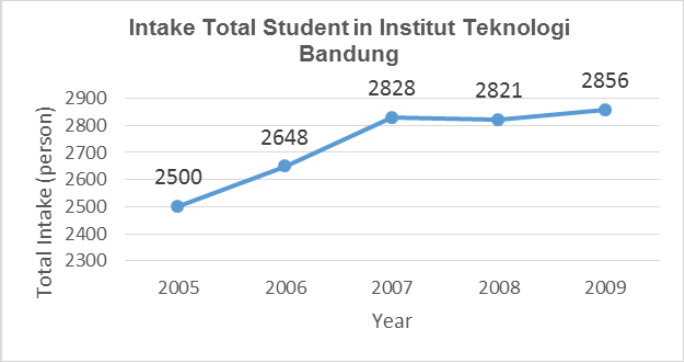



Fig 1. Line plot for total intake students of Institut Teknologi Bandung (2005-2009). It shows that in five years the intake total student was increasing regularly from 2005 till 2007, it slightly decreased in 2008 but it increased again in 2009.

Institut Teknologi Bandung (ITB) located in Bandung, the capital city of West Java Province, totally has 49 Undergraduate Programs (UP) from nine faculties, such as Mathematics and Natural Sciences, Technical Engineering, and three schools such as Pharmacy and Electrical engineering and Informatics. More than 20% UPs have been internationally certified by ASIIN, ABET, RSC, KAAB, AUN-QA, or ABEST21. For each year, ITB has around 3000 new students. Fig. 1 shows the growing population of ITB

students. The total intake of students in ITB increased quite rapidly from 2005 until 2007. Based on the number of students whose intake is getting higher, ITB realized that assessment and evaluation from its alumni are needed to improve the institution quality and achieve the level as a world class university.

There are more than 300 questions (variables) that respondents have to submit in their efforts to obtain feedback from alumni. These variables include biodata, assessment of educational quality and institutional facilities, and the relationship between university institutions and the world of work. Thus, the data obtained is very diverse, either in the form of quantitative or qualitative. Based on Hilbert's statement [3], the tracer study data could be said a Big data because it has a large data size with higher diversity.

Macatangay [4] researched about identification of the personal and professional characteristics, job placement of Computer Sciences graduates, and the school-related factors associated with their employment in Philippines University. This tracer study got 85% of the surveyed respondents those who are employed. Data analysis method used on the research is descriptive research design. Aquino *et.al* [5] worked on the similar method. Their main objective study was to trace the employment profile of the graduated bachelor in Secondary and Elementary Education, Philippines. Other research by Loquias [6] used descriptive research design to determine the employment status of graduates of BS Electronics Engineering from 1999 to 2011, Camarines Sur Polytechnic Colleges, Philippines. Sunchez and Diamante [7] studied the relevance and quality of programs offered by universities and labor market. Respondents of this tracer study are graduate students of Nursing at Cebu Lapulapu and Mandaue College. Those previous researches conducted their tracer study data in simplest way using descriptive statistics, meanwhile here, multivariate analysis is applied to obtain more comprehensive results.

Correspondence analysis (CA) is a multivariate analysis method that can be used to extract data. Benzecri [8] says that this method is a part of the nonparametric analysis that does not require assumption testing such as normality test, linearity, etc. Therefore, CA was chosen to analyze the tracer study data based on these advantages. It provides output in graphical display (correspondence maps) so that it is easier to understand. The correspondence map visualizes the relationship between two or more variables each consisting of two or more categories. The confidence circle on correspondence map was found by Beh [9] which leads to the easier interpretation of categories in CA. The confidence circle describes the area of a category with a certain degree of confidence in which the category is still considered significant. In general, if the confidence circle of a category passes through the point of origin on the correspondence map, then the category does not give a significant role to the corresponding variables.

This paper analyzes the relationship between two variables, i.e. batch of student intake in ITB and grade point cumulative (GPC) of respondents. Based on GPC, there are three classifications for student's achievement, i.e. Satisfied (St) for GPC ≤ 3.00, Very Satisfied (VS) for 3.00 < GPC ≤ 3.50, and Cum Laude (CL) for GPC > 3.50. Then, the batch students from academic year 2008 and 2009 are chosen as the categorical random variables. Both variables are interesting to be analyzed since some advices are maybe reached from the results, so that the university can make some improvement.

Furthermore, variable of GPC and waiting time to get first job are analyzed. Therefore, this study can give another advice to university for managing some programs or creating career guidance to prepare prospective graduates before getting a job. It is essential because it can provide significant support to an individual during their transition between levels and sectors of education and training systems and from school to adult and working life [10]. Then to obtain more comprehensive analysis using CA, the circle confidence is used.

The definition of some notations used in this paper and the utility of the confidence circle are described in Section 2. In Section 3, all about tracer study in ITB are described. Further description and processing of Tracer Study data using correspondence analysis are described in Section 4. The last section compares correspondence maps with different levels of confidence and conclusions obtained from the study.

## II. Correspondence Analysis

In this correspondence analysis, a new qualitative characteristic that can interpret the skin's visual characteristic is created. So, big data can be represented more simply. The qualitative characteristics specified should be mutually independent so that the independence test is performed on the contingency table. The principle that is done in a contingency table is the process of calculating, i.e. calculating the frequency of each category on the two selected variables. The contingency table in Fig. 2 illustrates two qualitative variables, such that random variables-1 consisting of $m$ row categories and random variables-2 consisting of $\ell$ column categories. The frequency of common events between the $i$-th row and $j$-th column is denoted as $n(i,j)$. While the $n_i$ and $n_j$ is respectively sum of frequencies for the $i$-th row category and $j$-th column category.

The next stage is to calculate the correspondence matrix **P,** $p_{ij}$ denotes the relative frequency value for each cross-combination of the $i$-th row and $j$-th column in the contingency table. The matrix **P** is written as follow

$$\mathbf{P} = \left(\frac{n_{ij}}{n_{..}}\right) = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1l} \\ p_{21} & p_{22} & \cdots & p_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{ml} \end{pmatrix}$$

where $n_{..} = \boldsymbol{N}$ denotes the total frequency in the contingency table.

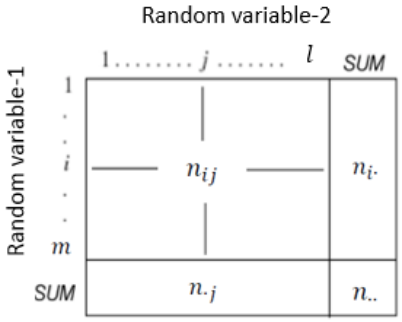



Fig. 2. The contingency table of two categorical variables. Variable $n_{ij}$ denotes the cross-combining frequency of the *i*-th row category and the *j*-th column category. The number of row categories is *m* and the column category is *l*.

To meet the assumption of independence, with the following hypothesis:

$$H_0: p_{ij} = p_{i.}p_{.j} \text{ vs } H_1: p_{ij} \neq p_{i.}p_{.j}$$

can be carried on by calculating the statistics test chi-squared on contingency table as follow

$$X^2 = \sum_i^m \sum_j^l \frac{\left(n_{ij} - \frac{n_{i.}n_{.j}}{n_{..}}\right)^2}{\frac{n_{i.}n_{.j}}{n_{..}}} = n_{..} \sum_i^m \sum_j^l \frac{(p_{ij} - p_{i.}p_{.j})^2}{p_{i.}p_{.j}} \quad (1)$$

For a level of significance $\alpha$, $H_0$ rejected if $X^2 > \chi^2_{\alpha,v}$ with degree of freedom $v = (m-1)(l-1)$.

Through the contingency table, there is a marginal opportunity distribution for the row category (row vector) $\boldsymbol{r}$ and column category (column vector) $\boldsymbol{c}$, respectively. Both vectors can be written as follows $\boldsymbol{r} = \begin{pmatrix} p_{1.} \\ p_{2.} \\ \vdots \\ p_{m.} \end{pmatrix}$ and $\boldsymbol{c} = \begin{pmatrix} p_{.1} \\ p_{.2} \\ \vdots \\ p_{.l} \end{pmatrix}$.

Coordinate determination on correspondence analysis is based on coordinate determination on Multi-Dimensional Scaling by Rancher [11]. Here, the residual matrix that is $\boldsymbol{P - rc^t}$ is standardized, so that the distance of each category can be compared. Consequently, CA does not require distribution assumptions. Standardization of the residual matrix is performed using (2).

$$\mathbf{S} = \boldsymbol{D_r^{-\frac{1}{2}}}(\boldsymbol{P - rc^t})\boldsymbol{D_c^{-\frac{1}{2}}} \quad (2)$$

with $s_{ij} = \frac{p_{ij} - p_{i.}p_{.j}}{\sqrt{p_{i.}p_{.j}}}$. The matrix $\boldsymbol{D_r^{-\frac{1}{2}}}$ and $\boldsymbol{D_c^{-\frac{1}{2}}}$ is defined respectively as

$$\boldsymbol{D_r^{-\frac{1}{2}}} = \begin{bmatrix} \frac{1}{\sqrt{r_1}} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{\sqrt{r_m}} \end{bmatrix}, \boldsymbol{D_c^{-\frac{1}{2}}} = \begin{bmatrix} \frac{1}{\sqrt{c_1}} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{\sqrt{c_l}} \end{bmatrix}$$

with $r_i$ is the sum from proportion for the $i$-th row and $c_j$ is sum from proportion for each the $j$-th column. In other words, they are marginal of probability density function for all categories. The existence of the matrix $\boldsymbol{D_r^{-\frac{1}{2}}}$ and $\boldsymbol{D_c^{-\frac{1}{2}}}$ is guaranteed because the values of $r_i$ and $c_j$ are always positive for $i = 1,2,..,m$ and $j = 1,2,\dots,l$. This paper used Singular Value Decomposition (SVD) on the matrix S as in (3) with the rank$(\mathbf{S}) = \min(m-1, l-1) \equiv k$ on dimention reduction method [12].

$$\mathbf{S} = \mathbf{UDV^t} \quad (3)$$

The matrix $\mathbf{D} = diag(\lambda_1, \lambda_2, \dots, \lambda_k)$ with $\lambda_t$ for $t = 1, 2, \dots, k$ are eigenvalues from $\mathbf{SS}^t$ matrix. $\mathbf{U}$ is a matrix whose columns are orthonormal eigenvectors corresponding to $\lambda_t$ where $\mathbf{U^tU = V^tV = I}$ [7]**.** The matrix $\mathbf{U}$ and $\mathbf{V}$ from SVD represent the main axis of the resulting CA map. Thus, the row category and column categories can be expressed as in (4). The coordinates of these row and column points can be presented simultaneously so that the characteristics of each category can be observed based on the proximity of the coordinate points.

$$\begin{aligned} X &= D_r^{-\frac{1}{2}} UD \\ Y &= D_r^{-\frac{1}{2}} V^t D \,. \end{aligned} \quad (4)$$

### *A. The value of diversity (Inertia)*

The value of inertia represents the proportion of data diversity described by a dimension/axis on the correspondence map so that the total inertia for all dimensions is 100%. Based on (1), the definition of total inertia according to Rancher [11] is as follows.

$$\begin{aligned} IT &= \frac{X^2}{n_{..}} \\ IT &= \sum_i^m \sum_j^l \frac{(p_{ij} - p_{i.}p_{.j})^2}{p_{i.}p_{.j}} \\ IT &= tr[D_r^{-1}(P - rc^t)D_c^{-1}(P - rc^t)^t] \\ IT &= tr[\mathbf{SS^t}] = \sum_{t=1}^k \lambda_t^2 \end{aligned} \quad (5)$$

Based on (5) the percentage contribution for each dimension $I_t, t = 1,2,\dots,k$ to total inertia can be calculated using (6).

$$I_t = \left(\frac{\lambda_t^2}{\sum_{t-1}^k \lambda_t^2}\right) \times 100\% \quad (6)$$

TABLE I
CONTINGENCY TABLE (A) AND CORRESPONDENCE TABLE (B) FROM BATCH (CATEGORY OF ROW) AND PREDICATE OF GRADUATION (CATEGORY OF COLUMN) AT FACULTY OF MATHEMATICS AND NATURAL SCIENCE.

| Batch | GPC | | | *SUM* |
|---|---|---|---|---|
| | St | VS | CL | |
| **2008** | 58 | 200 | 44 | ***302*** |
| **2009** | 10 | 6 | 2 | ***18*** |
| ***SUM*** | ***68*** | ***206*** | ***46*** | ***320*** |

**(a)**

| Batch | GPC | | | *SUM* |
|---|---|---|---|---|
| | St | VS | CL | |
| **2008** | 0,181 | 0,625 | 0,138 | ***0,944*** |
| **2009** | 0,031 | 0,019 | 0,006 | ***0,056*** |
| ***SUM*** | ***0,212*** | ***0,644*** | ***0,144*** | ***1,000*** |

**(b)**

As an illustration, data of Faculty of Mathematics and Natural Sciences data and GPC from two batches 2008 and 2009 are analyzed. The contingency and correspondence table is shown in Table Ia and Ib.

From Table Ia and Ib, the respondents from Batch 2008 is higher than Batch 2009. The trend of the different batch with GPC can be displayed in the correspondence map. By (1), (2), and (4) the result is given in Fig. 3.

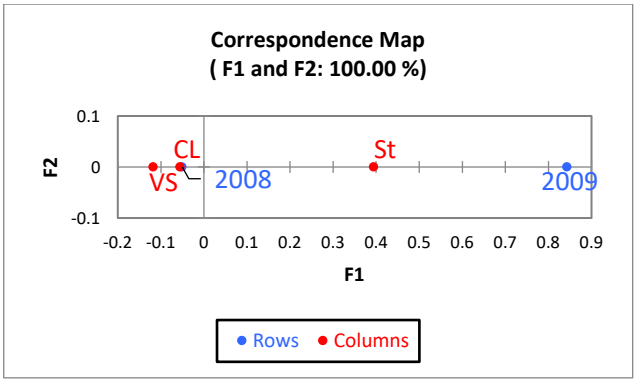



Fig. 3. Correspondence map of GPC vs batch for Faculty of Mathematics and Natural Sciences. It shows that alumni with GPC Very Satisfied and Cum Laude was dominated by alumni from Batch 2008, meanwhile Satisfied was nearer to Batch 2009.

Fig. 3 shows that alumni with academic achievement Very Satisfied and Cum Laude was dominated by alumni from Batch 2008. It happens logically since the number of respondents from Batch 2008 is more than Batch 2009. The eigenvalues from first dimension (F1 from Fig. 3) are 0,042 and 0. Furthermore, from the (6) the inertia value can be calculated. It is obvious that the inertia value for the first dimension is 100%. In other words, both the batch row category and the Predicate of Graduation column category can be mapped just on the first dimension.

### *B. Confidence circle*

In 2001 Beh demonstrated the confidence circle to simplify in interpreting the characteristics of a category through the corresponding graphical map of correspondence. He applied confidence circle on CA for analyzing drug and bean data. If a confidence circle from a row (or column) category passes through the origin, then the category contributes to the second independence hypothesis of random variables. Thus, the test statistic in (1) can be released as in (7).

$$X^2 = \boldsymbol{n} \times tr(\mathbf{D}^2) = tr(\boldsymbol{X^t D_r X}) = tr(\boldsymbol{Y^t D_c Y}) \qquad (7)$$

Radius of the confidence circle of a category can be calculated by (8) below. Notation of $\chi^2_{(\alpha,J-1)}$ is the value of the chi-square table for the level of significance $\alpha$ and degree of freedom $J-1$. The radius of this confidence circle is a random variable following the distribution of $t_{(\alpha,v_r)}$ with degree of freedom $v_r = \frac{6-(n_{..}\times p_{i.})}{n_{..}\times p_{i.}}$.

$$r_i = \sqrt{\frac{\chi^2_{(\alpha,J-1)}}{n_{..}\times p_{i.}}},\ r_j = \sqrt{\frac{\chi^2_{(\alpha,J-1)}}{n_{..}\times p_{.j}}} \qquad (8)$$

According to (8), the radius of the confidence circle depends on proportion of marginal for all category in contingency table. If the marginal proportion of a row (or column) category is greater than the circle confidence will be smaller.

## III. TRACER STUDY IN INSTITUT TEKNOLOGI BANDUNG

In 2010, research on tracer study was executed by a study team that held by Student Associate in ITB. The total of respondents that fill the questionnaire are increasing year by year. It can be seen in Fig. 4. The response rate in 2012 is 49%. It means that 49% of all alumni batch 2005 was filled in the questionnaire. Some previous researches, as mentioned before, obtained response rate 85% (on average). If we compare the response rate, ITB still has a low response. But in 2015, the response rate was reached 93%. It is an amazing result because high response rate on tracer study data very rarely.

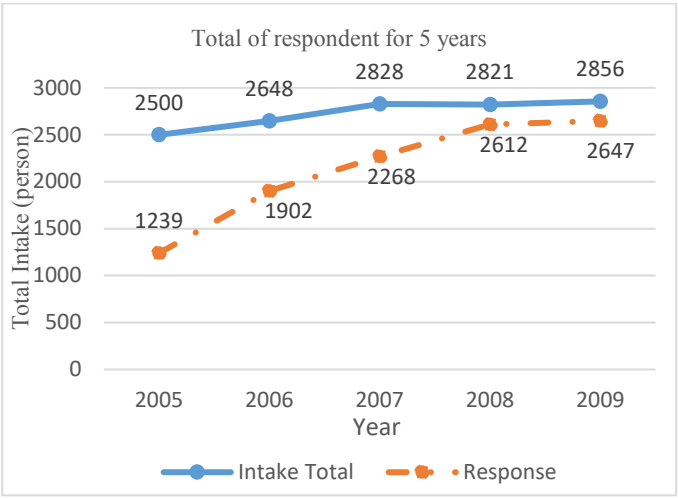



Fig. 4. Growth of total respondents in 5 years (2005-2009). Response rate in 2012 is 49%. It means that 49% of all alumni batch 2005 was filled in the questionnaire. in the last year (2015), the response rate reaches greater than 90%.

Since 2012, all of the alumni can fill the questionnaire online on the tracer study website. They can log in in Tracer Study website and fill the questionnaire everywhere and every time. Consequently, ITB has a high response rate year by year. All questions in the questionnaire were adapted from University Graduates Tracer Study Course (UNITRACE). However, the number of questions is raised every year according to the needs of the university. For example, Tracer study team gave 300 questions for the first batch (2010). For the next batch the team added or deleted some questions according to the evaluation last research. Usually, the team needs time to collect all data in 3 months.

The ITBs tracer study data can be called a big data because it has more than 300 variables and big sample size (Hilbert, 2016). The Tracer Study team in ITB used descriptive statistics method on their research. In the last report, the team compared the data from 3 different graduated years by graph. All reports about result of analysis tracer study data can be downloaded on the website.

## IV. Result and Discussion (Case Study)

The data of GPC alumni from Batch 2008 and 2009 was carried out in this paper. GPC is an important variable because it is one of the requirements in applying for a job. Total respondents for each batch are 2612 and 2647 persons respectively. However, there are some observations that fill in its GPC value zeros. This possibility occurs because of input errors while filling in an online questionnaire. The descriptive statistics of GPC data can be seen in Table II. It shows that the mean of Batch 2008 is higher than Batch 2009 with less standard deviation value. Both of them have negative skewnesses. Thus, the value of achievement index of both batches is piled up in high value. The histogram of GPC for each batch is also shown in Fig. 5 to support the summary.

TABLE II
DESCRIPTIVE STATISTICS OF GRADE POINT FROM ITB ALUMNI (BATCH 2008 AND 2009).

| Summary | 2008 | 2009 |
|---|---|---|
| Nbr. of observations | 2609 | 2646 |
| Minimum | 2,350 | 2,290 |
| Maximum | 3,970 | 3,990 |
| 1st Quartile | 3,060 | 3,070 |
| Median | 3,270 | 3,270 |
| 3rd Quartile | 3,480 | 3,480 |
| Mean | 3,270 | 3,269 |
| Interquartile range | 0,420 | 0,410 |
| Range | 1,620 | 1,700 |
| Standard deviation (n-1) | 0,288 | 0,295 |
| Skewness (Pearson) | -0,111 | -0,178 |
| Kurtosis (Pearson) | -0,321 | -0,186 |

Furthermore, Fig. 5 shows that the data have some lower outliers. But the outliers cannot be dropped because their proportion from full data is very small. The number of lower outliers on Batch 2009 (13 observations) is more than Batch 2008 (10 observations). It means that the lower GPC of Batch 2009 is less than batch 2008. According to these results, still a conclusion about willingness between GPC and difference of batch cannot be made, as a result, the CA should be done.

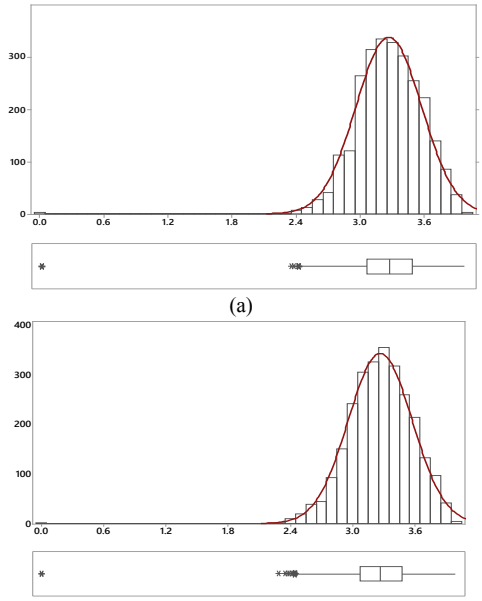



Fig. 5. Box-plot of grade point data from Batch 2008 (a) and 2009 (b). Horizontal and vertical axis in this figure is Grade Point Cumulative GPC value and frequency respectively. Both of them have outliers. The number of lower outliers on Batch 2008 is 10 respondents and on Batch 2009 is 13 respondents. It means that the lower GPC of Batch 2009 is less than Batch 2008.

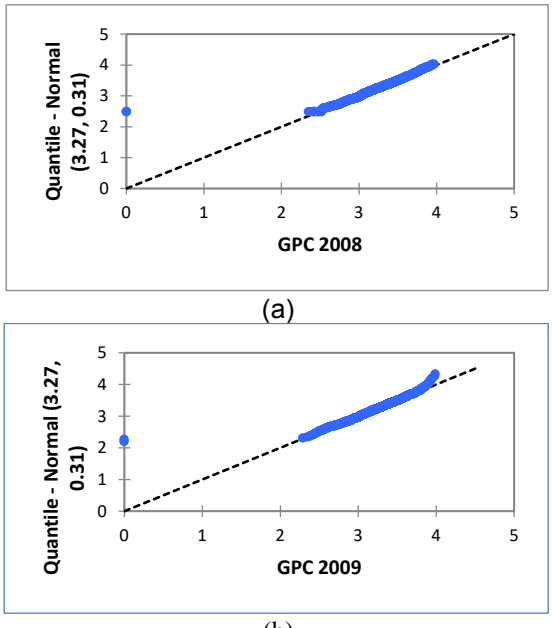



Fig. 6. Normal Q-Q Plot of grade point data from (a) batch of 2008 and (b) batch of 2009. The tail in figure (a) is closer than (b) to the diagonal line, but both batches are still concluded independence.

Before using CA, Chi-square test must be carried on to satisfy the constraint in CA that selected categorical variables must be independent. The statistic test value can be determined by (1), i.e. 0.3314. More, QQ Plot, in Fig. 6, shows that GPC value for both batches are independent because the tail of figure for each batch is close to diagonal.

Note that if it is far away from diagonal, the constrain is not satisfied. Based on the table then $H_0$ is not rejected for any significance level. This means that those two selected categorical variables have no significant effect. Hence, the CA method can be applied to this data.

After categorizing the data as described previously, obtained the contingency table was obtained as shown in Table III. Each value in the contingency table shows the cross-category combinations of row and column categories. Correspondence table of the data is shown in Table IIIb.

TABLE III
CONTINGENCY TABLE (a) AND CORRESPONDENCE TABLE (b) FROM BATCH (CATEGORY OF ROW) AND PREDICATE OF GRADUATION (CATEGORY OF COLUMN) FOR ALL FACULTIES.

| Batch | GPC | | | *SUM* |
|---|---|---|---|---|
| | St | VS | CL | |
| **2008** | 442 | 1575 | 592 | ***2609*** |
| **2009** | 460 | 1583 | 601 | ***2644*** |
| ***SUM*** | ***902*** | ***3158*** | ***1193*** | ***5253*** |

**(a)**

| Batch | GPC | | | *SUM* |
|---|---|---|---|---|
| | St | VS | CL | |
| **2008** | 0.084 | 0.300 | 0.113 | ***0.497*** |
| **2009** | 0.088 | 0.301 | 0.114 | ***0.503*** |
| ***SUM*** | ***0.172*** | ***0.601*** | ***0.227*** | ***1,000*** |

**(b)**

Table IIIb shows that the distribution of predicate categorical variable is not symmetric. St, VS, and CL graduation category have mass 0.172, 0.601, and 0.227, respectively. In contrast, this table shows that the distribution of row category, i.e. batches, is quite even. The frequency relative for batch 2008 and 2009 is 0.497 and 0.503, respectively.

Furthermore, a marginal opportunity distribution for the row category (row vector) and column category (column vector) is $\boldsymbol{r} = \begin{pmatrix} 0.497 \\ 0.503 \end{pmatrix}$, $\boldsymbol{c} = \begin{pmatrix} 0.172 \\ 0.601 \\ 0.227 \end{pmatrix}$ respectively. The standardized residual matrix **S** is

$$\mathbf{S} = \begin{pmatrix} -0.0039 & 0.0023 & -0.0003 \\ 0.0039 & 0.0023 & 0.0003 \end{pmatrix}$$

Then by SVD method, the matrix **U, D**, and **V** can be obtained. The results are

$$\mathbf{U} = \begin{pmatrix} -0.7095 & 0.7047 \\ 0.7047 & 0.7095 \end{pmatrix},$$

$$\mathbf{D} = \begin{pmatrix} 0.0064 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \text{and}$$

$$\mathbf{V} = \begin{pmatrix} 0.8627 & -0.2173 & -0.4566 \\ -0.5015 & -0.4853 & -0.7163 \\ 0.0658 & -0.8469 & 0.5277 \end{pmatrix}$$

From (4) the coordinate row category (in this case Batches), that is **X,** can be estimated i.e.

$$\mathbf{X} = \begin{pmatrix} -0.0064 & 0 & 0 \\ 0.0063 & 0 & 0 \end{pmatrix},$$

and the coordinate column category (in this case GPC), that is **Y,** i.e.

$$\mathbf{Y} = \begin{pmatrix} 0.0133 & 0 & 0 \\ -0.0018 & 0 & 0 \\ -0.0061 & 0 & 0 \end{pmatrix}$$

Fig. 7 is the graph of correspondence map without circle confident.

In the Fig. 7, it can be seen that batch 2008 has a tendency to graduate with Very Satisfied VS or Cum Laude CL. Whereas, for batch 2009 has a tendency to graduate with Satisfied St. Thus, there is an allegation that the 2008 class students are smarter than 2009 in accordance with its higher average value. Another possibility the lecturer who teaches then has a higher success rate in teaching. To get a better result, confidence circle was applied in this analysis.

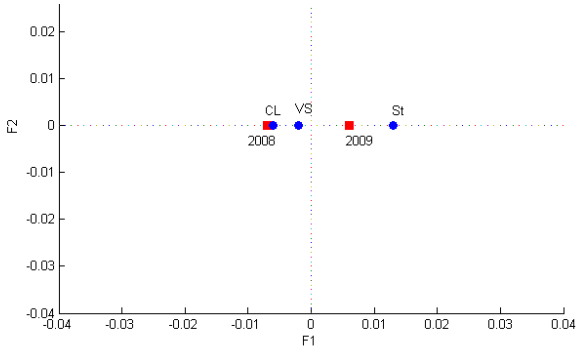



Fig. 7. Correspondence map without confidence circle. Horizontal and vertical axis is first and second dimensions, respectively. The blue one on the figure are coordinate point for column category (GPC) and the red one for row category (batch). It can be seen that the 2009 generation has a tendency to graduate with Satisfied (St) predicate.

According to (8) in Section 2, it has been seen that the large radius of the confidence interval is affected by the value of the chi-square table. Therefore, in Figure 8, it can be seen the influence of different region of confidence on the correspondence map. If the confidence area is increase, then the circle that is formed will be even greater. In Fig. 8 (a) that have a wider confidence region, the confidence circle for all categories always passes the point of origin. The same as the confidence circle for all categories always passed the point of origin though with a small confidence region of 40%. It means that for all confidence region, all categories did not have significant influence. We can say that batch of 2008 has a tendency to graduate with VS and CL predicate greater than batch 2009.

GPC data is influence waiting time to get a first job also. From the result before, batch 2008 has tendency to graduate with VS or CL. Practically, they have a short waiting time to get a first job after graduated because good GPC (VS or CL) is an important thing on the job requirement. In the questionnaire, there are two different questions about waiting time, i.e. before and after graduation. The alumni who seek the job before graduated do not use their GPC to apply, so waiting time after graduated data have chosen to be analyzed by CA.

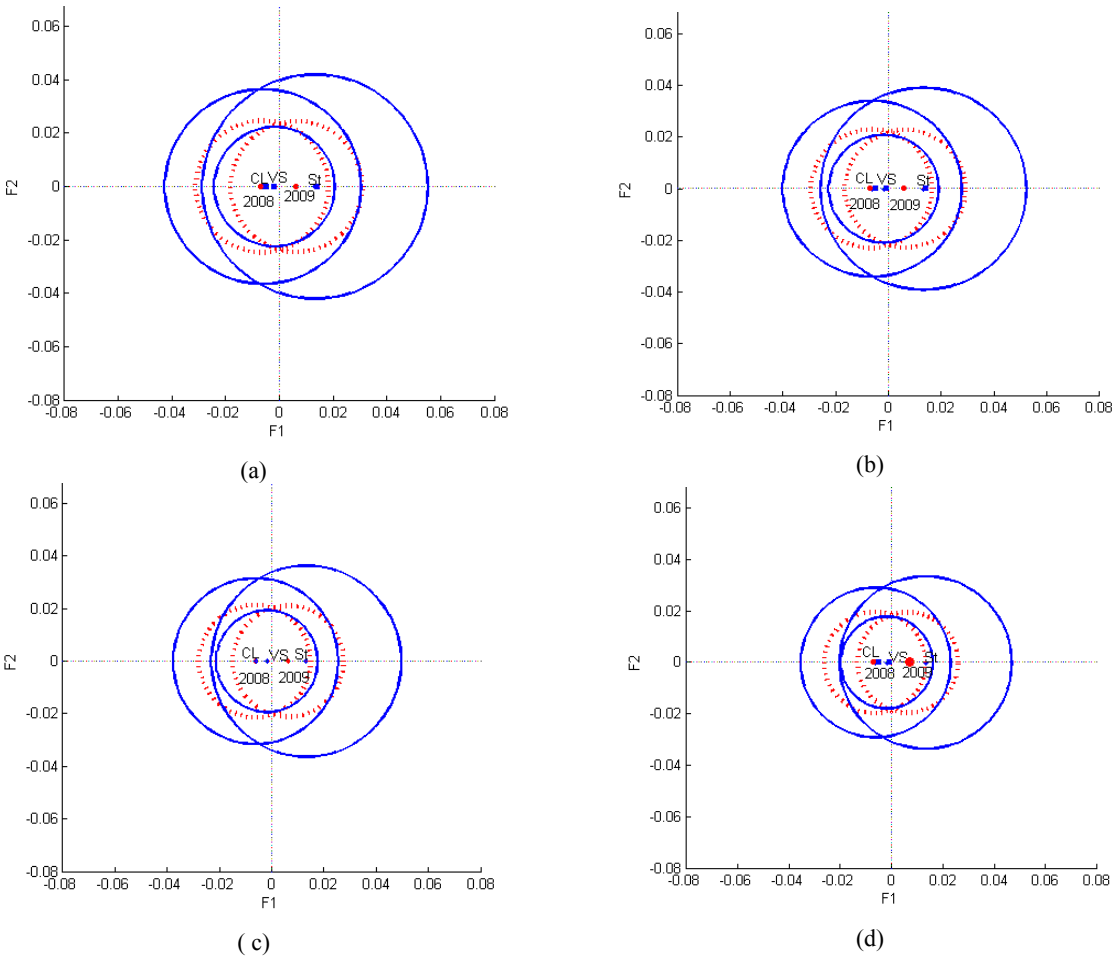



Fig. 8. Correspondence map with circle confidence. It can be seen that greater confidence region give greater radius of circle confidence (a) 55%, (b) 50%, (c) 45%, and (d) 40%. Blue and red circles are confidence circle for column and row categories. The confidence circle for all categories always passed the point of origin though with a small confidence region of 40%. It means that all categories did not have significant influence for all confidence region.

Table IV and Fig. 9 are statistics summary and graph from waiting time data. From Table IV we can see that the fastest and latest of waiting time is 1 month and 34 months (outlier in Fig. 7). Kurtosis from this data is 11,6. It means that the data is far from normal distribution and dominated with lower data. In other words, more than 70% of alumni have short waiting time. In other words, that the alumni disposed of get first job faster.

TABLE IV
DESCRIPTIVE STATISTICS OF WAITING TIME DATA FOR ALL ALUMNI IN BATCH 2008.

| Summary | 2008 | Summary | 2008 |
|---|---|---|---|
| Nbr. of obs. | 1345 | Mean | 4,3 |
| Minimum | 1 | Interquartile range | 4,0 |
| Maximum | 34 | Range | 33,0 |
| 1st Quartile | 2 | Standard deviation (n-1) | 18,1 |
| Median | 3 | Skewness (Pearson) | 2,8 |
| 3rd Quartile | 6 | Kurtosis (Pearson) | 11,6 |

CA was applied to see how far the tendency between waiting time and GPC. The unit for waiting time data is month. It is divided to two categories, Short (less than or equal to 6 months) and Long (greater than 6 months). The contingency table of both data (Batch 2008) can be seen in Table V.

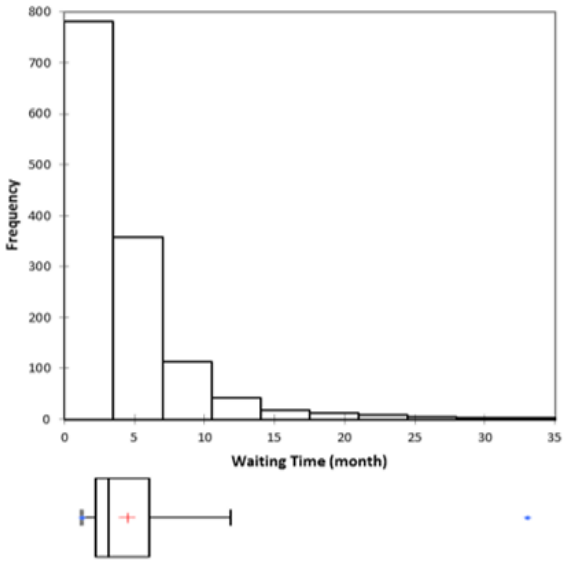



Fig. 9. Histogram and box plot of waiting time data. It shows that the data have an upper outlier and more than 70% it has lower value.

TABLE V
CONTINGENCY TABLE (a) AND CORRESPONDENCE TABLE (b) FROM WAITING TIME DATA (CATEGORY OF ROW) AND GPC (CATEGORY OF COLUMN) OF BATCH 2008. THE PROPORTION OF ALUMNI THAT WHO HAVE SHORT WAITING TIME TO GET FIRST JOB IS GREATER THAN WHO HAVE LONG WAITING TIME. IT IS MEAN THAT MORE THAN 50% FROM ALUMNI GET FIRST JOB FAST.

| Waiting Time | GPC | | | *SUM* |
|---|---|---|---|---|
| | St | VS | CL | |
| **Short** | 181 | 714 | 244 | ***1139*** |
| **Long** | 35 | 129 | 42 | ***206*** |
| ***SUM*** | ***216*** | ***843*** | ***286*** | ***1345*** |

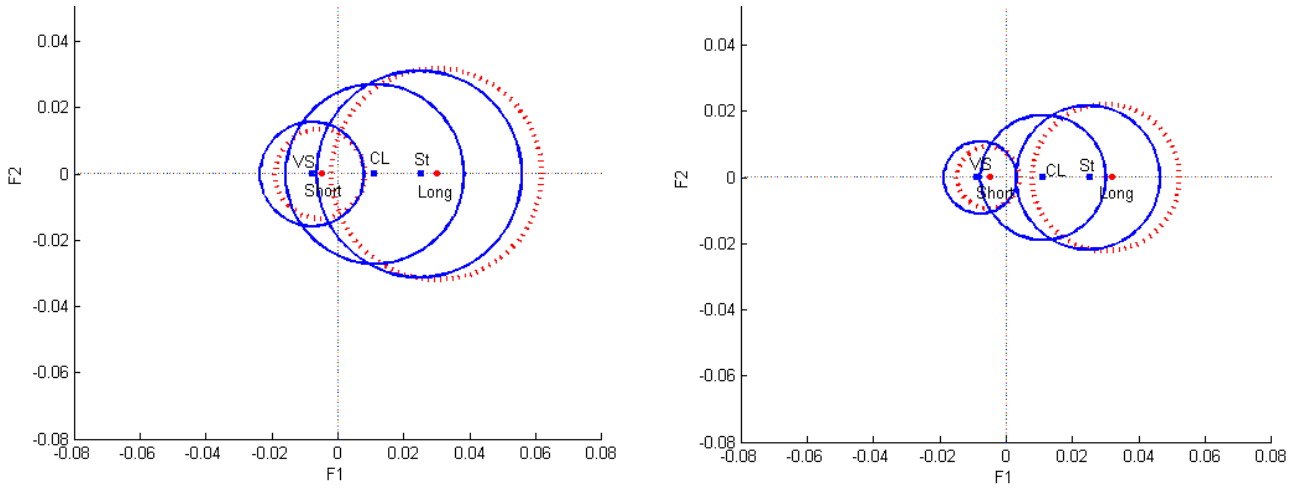
**(a)**

| Waiting Time | GPC | | | *SUM* |
|---|---|---|---|---|
| | St | VS | CL | |
| **Short** | 0,135 | 0,531 | 0,181 | ***0,847*** |
| **Long** | 0,026 | 0,096 | 0,031 | ***0,153*** |
| ***SUM*** | ***0,161*** | ***0,627*** | ***0,213*** | ***1,000*** |

**(b)**

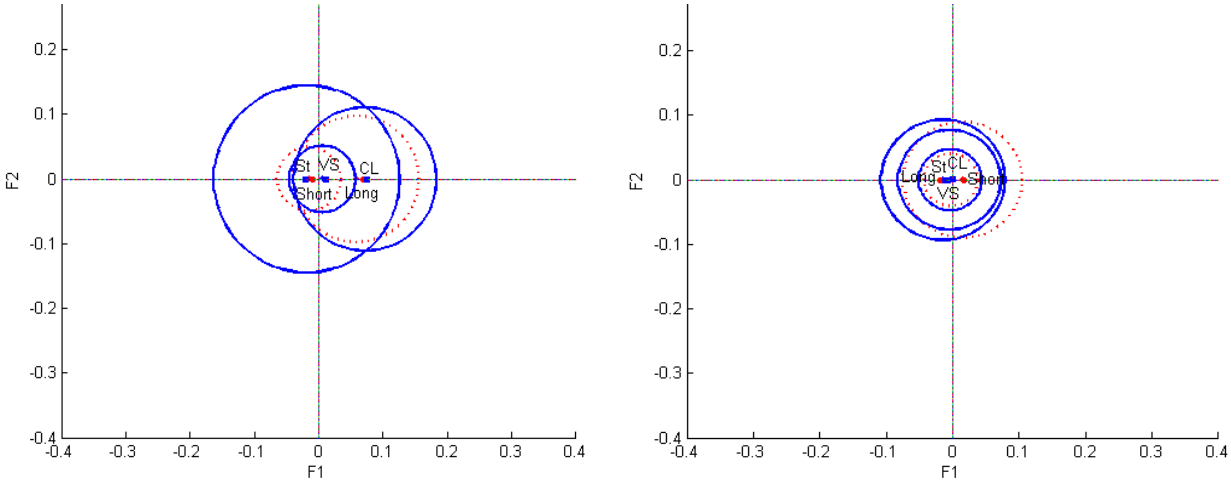
The proportion of alumni with short waiting time is higher than long waiting time (Table Vb). In addition, more than 50% of alumni who have short waiting time are alumni with VS predicate. Otherwise, the alumnus that just has St predicate tend to long waiting time. So, if the alumni get St predicate then they must be prepared with another skill to get first job faster. Corresponding map with confidence circle of both data can be observed in Fig. 10.

In Fig. 10a, the correspondence map with greater than 10% confidence region still cannot have significant influence. They have significant influence between satisfied (ST) of GPC and long waiting time with 5% confidence region. It can be seen in Fig. 10b that the confidence circle of them is not passed the origin. That analysis can be done for all faculties in ITB. Generally, almost each of faculty has similar behavior with CA above. But Faculty of Mining and Petroleum Engineering and School of Electrical Engineering and Informatics have different result of CA, which can be seen in Fig. 11.




(a) (b)

Fig. 10. Correspondence map with circle confidence from GPC and waiting time data of batch 2008. It can be seen that greater confidence region give greater radius of circle confidence (a) 10% and (b) 5%. The correspondence map with 10% confidence region still cannot have significant influence, but with 5% confidence region satisfied (ST) of GPC have significant influence with long waiting time.




(a) (b)

Fig. 11. Correspondence map with circle confidence from GPC and waiting time data of Batch 2008 for (a) Faculty of Mining and Petroleum Engineering and (b) School of Electrical Engineering and Informatics. It can be seen that greater confidence region give greater radius of circle confidence, around 10%. The

correspondence map with 10% confidence region still cannot give significant influence.

Both faculties have a similar tendency. All of the category falls in line, so for small confidence circle also does not have significance influence. Furthermore, the proportion of alumni with short waiting time is still higher for both faculty. It means that most of the alumni did not try too hard to get first job with different GPC. On both faculties, GPC did not have influence for alumni to get first job faster, especially on Faculty of Electric Engineering and Informatics. They have some skill (out of academic) to pass a possible requirement. Other than that, the current technological and information development provide more jobs also. It can be further analyzed on the suitability of the field of work.

## V. CONCLUSION

From the previous section, alumni from Batch 2008 have tendency to graduate with Very Satisfied and Cum Laude predicate compare to Batch 2009. It may happen because the lecturers who were teaching in that batch have excellent methods or systems. All faculties can do deeper research about that, to increase the quality of alumni. According to that condition, CA with confidence circle applied to analyze about tendency of GPC and waiting time to get first job data from batch 2008.

The confidence circle on the correspondent map for this data gives more information with very small confidence region. However, the conclusion of this analysis still can be taken. Alumni that graduated with very satisfied and cum laude have tendency to get first job faster (less than 6 months). Generally, almost all faculties have similar behavior, although there are some faculties did not have significant tendency between GPC and waiting time to get first job. Nevertheless, to achieve and maintain the best quality of alumni from year to year, the university should be prepared with some plans such as career guidance for students with less academic achievements.

## ACKNOWLEDGEMENT

All of the authors want to thank you to ITB tracer study team and Innovation Research Grant that was supported this research.

## REFERENCES

[1] H. Schomburg, *Handbook for Graduate Tracer Study: Centre for Research on Higher Education and Work*. Kessel: University of Kessel, 2003.

[2] Tracer study team of Collage of Business Education, *Tracer study report in CBE graduates in the labour market*, Tanzania, 2016.

[3] M. Hilbert, *The world's technological capacity store communicate, and compute information*. University of California, 2016.

[4] L. Macatangay, "Tracer study of BSCS graduates of lyceum of the Philippines University from 2004-2009", *Academic Research International*, Vol. 4, No. 5, 2015.

[5] A.B. Aquino, et.al, "Teacher education graduate tracer study from 2010 to 2014 in one state university in Batangas, Philipines", *Asia Pacific Journal of Multidisciplinary Research"*, vol. 3, No. 5, pp. 45-50, 2015.

[6] R.T. Loquias, "Employability of the Bachelor of Science in Electronics Engineering Graduates of Camarines Sur Polytechnic Colleges", *Asia Pacific Journal of Multidisciplinary Research,* Vol. 3, No. 4, pp. 33-40, 2015.

[7] M.P.R. Sanchez and V.A.M. Diamante, "Graduate tracer study of the collage of nursing", *The Malaysian Journal of Nursing*, Vol. 8, No. 3, 2017.

[8] J.P. Benzecri, *Correspondence Analysis Handbook*, Marcel Dekker Inc, New York, USA, 1992.

[9] E.J. Beh, "Confidence circles for correspondence analysis using orthogonal polynomials", *Journal of Applied Mathematics and Decision Science*, vol. 5, no. 1, pp. 35-45, 2001.

[10] M. Debono, A. Debono, A., and N. Caruana, *Career Outcomes of Graduates 2004. A Career Guidance Tool*, Euroguidance Malta, Malta, 2005.

[11] A.C. Rancher, *Method of Multivariate Analysis (second edition),* John Wiley and Sons, Canada, 2002.

[12] E.J. Beh, "Simple correspondence analysis: a bibliographic review", *International Statistical Review*, vol. 72, No. 2**,** pp. 257-284, 2007.

Internetworking Indonesia Journal
ISSN: 1942-9703
www.InternetworkingIndonesia.org