Employment Impact Analysis using Variance Analysis based on Student Data

Eri DOMOTOa *, Koji OKUHARAb

aDepartment of Media Business, Faculty of Economics, Hiroshima University of Economics, 5-37-1, Gion, Asaminami, Hiroshima-shi, Hiroshima, 731-0192 Japan
bDepartment of Electrical and Computer Engineering, Graduate School of Engineering, Toyama Prefectural University 5180 Kurokawa, Imizu, Toyama 939-0398, Japan

Abstract

One of the goals of university students is to get a job at a good company. Students graduate from high school and enter the university using various entrance examinations. Then, receiving a lecture at a university, acquiring a unit, assigning to a seminar and preparing a graduation thesis. After graduating from university, most students find employment in companies. In this way, it is possible to acquire various data in a series of flows from high school graduation to employment. This study analyzes how a set of factors such as deviation value of the senior high school, type of entrance examination, achievement of the University, seminar advisor, and the employment place data can influence this process. Moreover, the paper considers the analysis of variance for the same purpose.

Keywords: Factor Analysis, Employment support, Education Data Analysis, Analysis of Variance, Data Mining

1. Introduction

The purpose of this study is to clarify the influence of various factors of university life on employment. Factors influencing employment include faculty and department, high school deviation value, types of entrance examination, club, project, GPA (Grade Point Average), scholarship, advisor and so on. At university, making students enrolled in various exams into society is regarded as one index in university educational institutions, and employment support is important. It is thought that various factors of university life, such as grades and club activities, will lead to employment [1-4]. GPA is a grading evaluation value of a student calculated by a specific method from the performance of each subject, or a grading evaluation method thereof. It is commonly used in Western universities and high schools, and it is an index to measure academic ability, such as when you study abroad. In this study, we analyze the deviation value of the high school before entering university, the type of entrance examination, the GPA during university studies, and how the teacher in charge of seminars influences the process for employment. Further, we analyze the effect of each factor by using two-way ANOVA [5, 6] with no repetition.

2. Classification of Education Data

In this study, we analyze using various data from entering university until graduation. Approximately 50 students who graduated in 4 years since enrolling were targeted. For the student data, high school deviation value, types of entrance examination, GPA, adviser, place of employment was used. The high school deviation value is classified as shown in Table 1.

Table 1. Classification of high school deviation value

| Deviation | Classification |
|-----------|----------------|
| 55-64.9   | A              |
| 50-54.9   | B              |
| 45-49.9   | C              |
| 40-44.9   | D              |
| 0.1-39.9  | E              |
| unknown   | N              |

The types of entrance examination is classified as shown in Table 2. General method of calculating GPA is the grade for each subject evaluated in five steps, by applying an assessment of 4-0 points, and calculates the grade average value per one unit. Marks for Grade (point) is as follows. AA (90–100 point) is 4 point, A (80–89 point) is 3 point, B (70–79 point) is 2 point, C

*Corresponding author: er-domo@hue.ac.jp

Received: March 31, 2017
Accepted: November 21, 2017
Table 2. Classification of the type of entrance examinations

| Types of entrance examinations                                           | Classification |
|-------------------------------------------------------------------------|-----------------|
| General Entrance Exam (Term 1)                                          | 1               |
| General Entrance Exam (Term 2)                                          |                 |
| Entrance Exam for General Public Advertisement Recommended              | 2               |
| Special Designated School Recommendation Entrance Examination          | 3               |
| Designated School Recommendation Entrance Examination                  | 4               |
| Designated Club Recommendation Entrance Examination                    |                 |
| Entrance Exam for Special Recommended (Sports Field)                   | 5               |
| Entrance Exam for Special Recommended (Specific Qualification Field)    |                 |
| Admissions Office Entrance Exams                                       |                 |
| Entrance Exam for International Students                                |                 |
| University Center Examinations (Term 1)                                 | 6               |

(60–69 point) is 1 point, F (Fail) (59 point or less, evaluation impossible, pending) is 0 point. (Units of AA * 4 point + units of A * 3 point + units of B * 2 + units of C * 1 point) / Sum of units of subject to be studied = GPA point. Classification of GPA is: A: 3.00-4.00, B: 2.50-2.99, C: 2.00-2.49, D: 1.50-1.99, E: 0.00-1.49.

Adviser is a faculty who instructs graduation thesis. The adviser categorized as three, professor, associate professor, assistant professor.

The employment place is classified into three. Many students from class1 wanted to go to 100 specific companies, class2 is public servant, class3 wanted to go to companies not included in the specific 100 companies. Class4 chose other than employment.

3. Factor analysis of influence on employment

Here, we analyze which factors are affecting employment for the four employment place. There are 738 target students, the number of people to be employed is class 1: 54, class 2: 28, class 3: 545, class 4: 111. As a factor, compare high school deviation value, GPA, types of entrance examination, club / projects, and scholarships.

Figure 1 shows the average of high school deviation value and GPA. In the high school deviation value, it is understood that the student who is employed by public servant of class 2 is very high. The reason is that the content of the public servant exam is up to high school, so it is easy to pass the public servant examination if you are out of a high school with a high deviation value. Looking at the GPA, it turns out that students getting employed in class 1 are high. It is common in high school deviation value and GPA that students of class 4 have low high school deviation value and GPA.

Figure 2 shows the classification of employment place by types of entrance examinations. To Class 1, there are a lot of students who entered by using the entrance examination of the university center examinations. In addition, it can be seen that there are many students who use entrance exam for general public advertisement recommended and designated school recommendation entrance examination for class 2. This shows that the probability that a student with good grades at high school will be a public servant is high.
Figure 3 Classification of employment place by club / projects

Figure 3 shows the classification of employment place by club / project. It turns out that the student belonging to the project has a high proportion of class 1. This is considered to be actively acting in university life. Students in class 2 know that they belong somewhat. On the contrary, students in class 4 find that there are many students who do not belong to club activities, projects, etc. during university life.

Figure 4 Classification of employment place by get or not get scholarship

Figure 4 shows the classification of employment place by get or not get scholarship. From this figure, it is understood that students who are employed by companies, public servants, etc. have a scholarship. Students in class 4 have a small percentage of scholarships. Furthermore, among the class 2 students, the percentage of students who are obtaining scholarships is large.

From the above, the following can be said.

Class 1:
- High GPA
- There was a high proportion of entrance using the university center examinations.
- There are a lot of students belonging to the project.
- There are a lot of students who get a scholarships.

Class 2:
- High school deviation value is high.
- There is a high proportion of students who used the entrance exam for general public advertisement recommended and designated school recommendation entrance examination.
- There are many proportions belonging to club / projects.
- There are a lot of students who get a scholarship.

Class 3:
- There are many average students.

Class 4:
- High school deviation value and GPA are both low.
- The proportion of students belonging to club / projects is low.
- There are few students who have a scholarship.

4. Analysis of Variance (ANOVA)

4.1 Outline of ANOVA

An element having an effect on the observed value is called a factor. To make a plurality of conditions by changing the values of the factors referred to experimental research studies examining the value of the observations for each condition. This condition we called a level. First of interest in experimental studies, whether the difference in the level of influence for the observed value, that is, whether or not there is an effect of factors. If the number of factors taken up is one, then it is considered to be a one-way model, while a two-way model consists of two factors. Sometimes multiple layout models with three or more factors are used.

In this study we use the two-way model.

4.2 Outline of Two-Way ANOVA

The ANOVA model factors in two affecting the observed values are called a two-way layout. Two factors are A and B. J is the number of levels of A, K is the number of levels of B. r number of observations y_{ijk} (i = 1, 2, ..., r; j = 1, 2, ..., J; k = 1, 2, ..., K) per combination (A_j, B_k) of each level is obtained.

\[ y_{ijk} = \mu_{jk} + \epsilon_{ijk} \sim N(0, \sigma^2) \]  \hspace{1cm} (1)

Observations set the number of observations for each level combination (A_j, B_k). \( \epsilon_{ijk} \) is assumed to independently follow the same normal distribution \( N(0, \sigma^2) \) each other. \( y_{ijr} \) is an independent all. The following equation holds for each combination (A_j, B_k) of each level.

\[ y_{ijr} ~ N(\mu_{jk}, \sigma^2) \]  \hspace{1cm} (2)

By setting the following formula per combination (A_j, B_k) of each level, \( y_{ij} \), \( \bar{y}_{jk} \), etc. \( y_{ij} \) is independently from each other, \( y_{jk} \sim N(\mu_{jk}, \sigma^2/r) \) holds (j = 1, 2, ..., J; k = 1, 2, ..., K).
It is decomposed into overall average and the factors of effect the population mean $\mu_{jk}$.

Overall average
$$\mu = \frac{1}{JK} \sum_{j=1}^{J} \sum_{k=1}^{K} \mu_{jk}$$
(3)

Line average
$$\mu_j = \frac{1}{K} \sum_{k=1}^{K} \mu_{jk}$$
(4)

Column average
$$\mu_k = \frac{1}{J} \sum_{j=1}^{J} \mu_{jk}$$
(5)

The $j$ level effects of factors $A$, the effects of the $k$ levels of factor $B$, respectively it is defined by the following equation.
$$\alpha_j = \mu_j - \mu, \quad \beta_k = \mu_k - \mu$$
$$\alpha_j (j = 1, 2, \cdots, J; k = 1, 2, \cdots, K)$$
(6)

At this time, the following identity is established.
$$\mu_{jk} = \mu + (\mu_j - \mu) + (\mu_k - \mu) + (\mu_{jk} - \mu_j - \mu_k + \mu)$$
$$= \mu + \alpha_j + \beta_k + (\mu_{jk} - \mu_j - \mu_k + \mu)$$
(7)

The last term of the second line the right side, in the superposition of effects $\alpha_j$ factors $A$ and effect $\beta_k$ factors $B$ is the effect that cannot be explained.

It can be interpreted as a specific effect on the level combination $(A_j, B_k)$. This is called interaction effect, as defined by $(\alpha \beta)_{jk}$ in the following expression.
$$(\alpha \beta)_{jk} = \mu_{jk} - \mu_j - \mu_k + \mu$$
(8)

$$\mu_{jk} = \mu + \alpha_j + \beta_k + (\alpha \beta)_{jk}$$
(9)

Effects $\alpha_j, \beta_k$ are said to be the main effects. The model assumes a structure consisting of a model (8)(9) that two-way analysis of variance model. Each expression of the following is true from the definition.
$$\sum_{j=1}^{J} \alpha_j = 0, \sum_{k=1}^{K} \beta_k = 0$$
$$\sum_{j=1}^{J} (\alpha \beta)_{jk} = 0, \sum_{k=1}^{K} (\alpha \beta)_{jk} = 0$$
(10)

Two-way ANOVA tests hypotheses about the effects of factors $A$ and $B$, and their interaction on the response variable $y$. The hypotheses about the equality of the mean response for groups of row factor $A$ are
$$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_j$$
$$H_1: \text{at least one } \alpha_j \text{ is different, } j = 1, 2, \cdots, I$$
(11)

The hypotheses about the equality of the mean response for groups of column factor $B$ are
$$H_0: \beta_1 = \beta_2 = \cdots = \beta_j$$
$$H_1: \text{at least one } \beta_j \text{ is different, } j = 1, 2, \cdots, J$$
(12)

The hypotheses about the interaction of the column and row factors are
$$H_0: (\alpha \beta)_{ij} = 0$$
$$H_1: \text{at least one } (\alpha \beta)_{ij} \neq 0$$
(13)

5. Numerical Example

Here, we analyzed factors influencing place of employment by using analysis of variance based on 48 student data. Table 3 shows the number of people who apply to each factor and result of analysis of variance. From the analysis of variance table it is clear whether each of the factors affecting the number of people.

For example, Table 3-1(b) shows whether the types of entrance examinations and the high school deviation value influence the number of students.

Method of determining variance analysis results are as follows. Compare the observed variance ratio (F-value) and F boundary value (FB-value) and make the following determination.

1) If F-value $\geq$ FB-value, it is considered statistically significant at a significance level of 5% and it is considered to be effective.

2) If F-value $< FB-value$, it is not significant at a significance level of 5% and cannot be said to be effective.

3) The P-value is the smallest probability that the observed variance ratio is significant.

- If P-value $\leq 0.01$, it is significant at a significance level of 1%.
- If $0.01 < \text{P-value} \leq 0.05$, it is significant at a significance level of 5%.
- If P-value $> 0.05$, it cannot be considered significant.

Table 3-1(a) Number of people between High school deviation value and Types of entrance examinations

| Types of entrance examinations | High school deviation value |
|-------------------------------|-----------------------------|
|                               | A  | B  | C  | D  | E  | N  | Total |
|-------------------------------|----|----|----|----|----|----|-------|
|                               | 1  | 2  | 1  | 2  | 1  | 0  | 10    |
| 2                             | 0  | 2  | 2  | 3  | 0  | 0  | 7     |
| 3                             | 1  | 4  | 1  | 0  | 0  | 6  |       |
| 4                             | 1  | 1  | 5  | 7  | 1  | 16 |       |
| 5                             | 0  | 1  | 3  | 0  | 1  | 5  |       |
| 6                             | 0  | 4  | 0  | 0  | 0  | 4  |       |
| Total                         | 6  | 14 | 12 | 12 | 2  | 2  | 48    |

Table 3-1(b) Results of variance analysis of High school deviation value and Types of entrance examinations

| Source                          | SS  | DF | SD   | F    | P    | FB   | FB  |
|--------------------------------|-----|----|------|------|------|------|-----|
| Types of entrance examinations | 16.333 | 5  | 3.267 | 1.324 | 0.286 | 2.603 |
| High school deviation value    | 24  | 5  | 4.8  | 1.946 | 0.122 | 2.603 |
| Error                          | 61.667 | 25 | 2.467 |       |      |      |     |
| Total                          | 102 | 35 |      |       |      |      |     |

From Table 3-1(a), it can be seen that many students using high school deviation values of 50 or more use the general entrance examination (term 1, 2) and university center examinations (term 1). From Table 3-1(b), in types of entrance examinations, 1.32432 (F-value) $< 2.60299$ (FB-value), and 0.28614 (P-value) $> 0.05$, it is not significant at a significance level of 5% and cannot be said to be effective. As for high school deviation value, 1.94595 (F-value) $< 2.60299$ (FB-value), and 0.12223 (P-value) $> 0.05$, it is not significant at a significance level of 5% and cannot be said to be effective.
Table 3-2(a) Number of people between High school deviation value and GPA

| GPA     | A | B | C | D | E | N | Total |
|---------|---|---|---|---|---|---|-------|
| 1-4     | 4 | 3 | 3 | 6 | 1 | 0 | 17   |
| 2-5     | 0 | 2 | 5 | 1 | 0 | 1 | 9    |
| 3-6     | 1 | 5 | 0 | 2 | 0 | 0 | 8    |
| 4-7     | 1 | 3 | 3 | 2 | 1 | 1 | 11   |
| 5-8     | 0 | 1 | 1 | 1 | 0 | 0 | 3    |
| Total   | 6 | 14| 12| 12| 2 | 2 | 48   |

Table 3-2(b) Results of variance analysis of High school deviation value and GPA

| Source                      | SS  | DF | SD  | F   | P   | FB  |
|-----------------------------|-----|----|-----|-----|-----|-----|
| GPA                         | 17.2| 4  | 4.3 | 2.312| 0.093| 2.866|
| High school deviation value | 28.8| 5  | 5.76| 3.097| 0.031| 2.711|
| Error                       | 37.2| 20 | 1.86|     |     |     |
| Total                       | 83.2| 29 |     |     |     |     |

From Table 3-2(a), Student high school deviation value is high, relatively GPA also high. From Table 3-2(b), in GPA, 2.31183 (F-value) < 2.86608 (FB-value), and 0.09305 (P-value) > 0.05, it is not significant at a significance level of 5% and cannot be said to be effective. However high school deviation value, 3.09677 (F-value) > 2.71089 (FB-value), and 0.03136 (P-value) ≤ 0.05, it is considered statistically significant at a significance level of 5% and it is considered to be effective.

Table 3-3(a) Number of people between Types of entrance examinations and GPA

| GPA     | 1 | 2 | 3 | 4 | 5 | 6 | Total |
|---------|---|---|---|---|---|---|-------|
| 1-3     | 6 | 3 | 2 | 5 | 0 | 1 | 17    |
| 2-5     | 1 | 1 | 0 | 5 | 1 | 1 | 9     |
| 3-6     | 1 | 2 | 0 | 2 | 1 | 2 | 8     |
| 4-7     | 1 | 2 | 2 | 3 | 3 | 0 | 11    |
| 5-8     | 0 | 0 | 1 | 1 | 0 | 1 | 3     |
| Total   | 9 | 8 | 5 | 16| 5 | 5 | 48    |

Table 3-3(b) Results of variance analysis of Types of entrance examinations and GPA

| Source                      | SS  | DF | SD  | F   | P   | FB  |
|-----------------------------|-----|----|-----|-----|-----|-----|
| GPA                         | 17.2| 4  | 4.3 | 2.416| 0.083| 2.866|
| Types of entrance examinations| 18.4| 5  | 3.68| 2.067| 0.112| 2.711|
| Error                       | 35.6| 20 | 1.78|     |     |     |
| Total                       | 71.2| 29 |     |     |     |     |

From Table 3-3(a), Students enrolled in general entrance examination (term 1, 2) and designated School Recommendation Entrance Examination have high GPA. From Table 3-3(b), in GPA, 2.41573 (F-value) < 2.86608 (FB-value), and 0.08267 (P-value) > 0.05, it is not significant at a significance level of 5% and cannot be said to be effective. As for types of entrance examinations, 2.41573 (F-value) < 2.86608 (FB-value), and 0.08267 (P-value) > 0.05, it is not significant at a significance level of 5% and cannot be said to be effective.

Table 3-4(a) Number of people between High school deviation value and Adviser

| GPA     | A | B | C | D | E | N | Total |
|---------|---|---|---|---|---|---|-------|
| 1-3     | 4 | 3 | 3 | 6 | 1 | 0 | 17   |
| 2-5     | 0 | 2 | 5 | 1 | 0 | 1 | 9    |
| 3-6     | 1 | 5 | 0 | 2 | 0 | 0 | 8    |
| 4-7     | 1 | 3 | 3 | 2 | 1 | 1 | 11   |
| 5-8     | 0 | 1 | 1 | 1 | 0 | 0 | 3    |
| Total   | 6 | 14| 12| 12| 2 | 2 | 48   |

Table 3-4(b) Results of variance analysis of High school deviation value and Adviser

| Source                      | SS  | DF | SD  | F   | P   | FB  |
|-----------------------------|-----|----|-----|-----|-----|-----|
| Adviser                     | 25  | 3  | 8.333| 5   | 0.013| 3.287|
| High school deviation value | 36  | 5  | 7.2  | 4.32| 0.012| 2.901|
| Error                       | 35.6| 20 | 1.78|     |     |     |
| Total                       | 86  | 23 |     |     |     |     |

From Table 3-4(a), High school deviation value is high students, choosing a seminar professor. From Table 3-4(b), in adviser, 5 (F-value)> 3.28738 (FB-value), and 0.01337 (P-value) ≤ 0.05, it is considered statistically significant at a significance level of 5% and it is considered to be effective. As for high school deviation value, 4.32 (F-value) > 2.90129 (FB-value), and 0.01236 (P-value) ≤ 0.05, it is considered statistically significant at a significance level of 5% and it is considered to be effective.

Table 3-5(a) Number of people between GPA and Adviser

| GPA     | 1 | 2 | 3 | 4 | 5 | Total |
|---------|---|---|---|---|---|-------|
| 1-4     | 8 | 7 | 1 | 1 | 1 | 17    |
| 2-5     | 4 | 2 | 4 | 1 | 1 | 17    |
| 3-6     | 5 | 0 | 2 | 3 | 2 | 12    |
| 4-7     | 0 | 0 | 1 | 1 | 0 | 2     |
| Total   | 17| 9 | 8 | 11| 3 | 48    |

Table 3-5(b) Results of variance analysis of GPA and Adviser

| Source | SS  | DF | SD  | F   | P   | FB  |
|---------|-----|----|-----|-----|-----|-----|
| Adviser | 30  | 3  | 1.967| 0.173| 3.490|
| GPA     | 25.8| 4  | 6.45 | 1.269| 0.335| 3.259|
| Error   | 61  | 12 | 5.083|     |     |     |
| Total   | 116.8| 19 |      |     |     |     |

From Table 3-5(a), GPA is high students it can be seen that choosing a seminar professor. From Table 3-5(b), in...
adviser, 1.96721 (F-value) < 3.49029 (FB-value), and 0.17279 (P-value) > 0.05, it is not significant at a significance level of 5% and cannot be said to be effective. As for GPA, 1.26885 (F-value) < 3.25917 (FB-value), and 0.33530 (P-value) > 0.05, it is not significant at a significance level of 5% and cannot be said to be effective. As for GPA, 1.26885 (F-value) < 3.25917 (FB-value), and 0.33530 (P-value) > 0.05, it is not significant at a significance level of 5% and cannot be said to be effective.

Table 3-6(a) Number of people between Types of entrance examinations and Adviser

| Advisor | 1 | 2 | 3 | 4 | 5 | 6 | Total |
|---------|---|---|---|---|---|---|-------|
| 1       | 4 | 2 | 1 | 7 | 1 | 2 |  17  |
| 2       | 3 | 6 | 2 | 5 | 1 | 0 |  17  |
| 3       | 2 | 0 | 2 | 3 | 2 | 3 |  12  |
| 4       | 0 | 0 | 0 | 1 | 1 | 0 |   2  |
| Total   | 9 | 8 | 5 | 16| 5 | 5 | 48   |

Table 3-6(b) Results of variance analysis

| Source          | SS  | DF | SD  | F    | P     | FB  |
|-----------------|-----|----|-----|------|-------|-----|
| Advisor         | 25  | 3  | 8.333| 3.289| 0.050 | 3.287|
| Types of entrance examinations | 23  | 5  | 4.6  | 1.816| 0.170 | 2.901|
| Error           | 38  | 15 | 2.533|      |       |     |
| Total           | 86  | 23 |     |      |       |     |

From Table 3-6(a), Student high school deviation value is high, relatively GPA also high. From Table 3-6(b), in adviser, 3.28947 (F-value) > 3.28738 (FB-value), and 0.04991 (P-value) ≤ 0.05, it is considered statistically significant at a significance level of 5% and it is considered to be effective. However types of entrance examinations, 1.81579 (F-value) < 2.90129 (FB-value), and 0.17019 (P-value) > 0.05, it is not significant at a significance level of 5% and cannot be said to be effective.

6. Conclusion

In this study, we analyzed what kind of factor affects employment by using data from entering university until getting a job. Specifically, data mining was conducted for 48 students using high school deviation value, types of entrance examination, GPA, adviser, and place of employment data. We also analyzed which factors are affecting by using analysis of variance. In this study, because it was the analysis that target a small number of people, the future want to do the analysis to target more large number of people. Also, we would like to clarify the factors affecting employment with various analysis methods.

Acknowledgments

This work is supported by JSPS KAKENHI Grant Number 25350309.

References

[1] Y. Omae, K. Nakahira, Y. Tsuchiya, R. Shukui, T. Mitsui and H. Takahashi: Method to Detect Change of Motivation to Enroll in University by Survey of Career Perceptions. Japan Society for Fuzzy Theory and Intelligent Informatics, 27(5), (2015), 743-756.
[2] Y. Fujihara, H. Ohnishi and H. Kato: The Development of the Method to Revise an Evaluated Score in Considering of the Evaluator's Characteristic for a Peer Evaluation with Many Missing Responses (Application and Aggregation of Learning Objects and Learning Data). Japan Journal of Educational Technology, 31(3), (2007), 373-381.
[3] F. Watanabe and C. Kogo: Causal Relationship among Questionnaire Variables Investigating Learners' e-Learning Orientation in Massive Online Courses. Japan Journal of Educational Technology, (2017).
[4] R. Murai, T. Hayashi and R. Yaegashi: Study of modeling for intellectual property education using ICT. IEICE Technical Report, 113(482), (2014), 205-208.
[5] H. Suzuki, K. Techavorasinsakul and T. Enkawa: Conjoint Analysis Incorporating the Outer-Array and Its Application to Product Planning. Journal of Japan Industrial Management Association, 47(4), (1996), 257-264.
[6] M. Kera, M. Nakaya, T. Umemoto, M. Nakashishi and K. Yanagisawa: The Effect of Utility Value Intervention on Japanese Undergraduate Students’ Motivation in Educational Psychology Classes. Japan Journal of Educational Technology, 40, (2017), 57-60