**데이터 애널리틱스 1차 과제**

* **학번: 201402167**
* **성명: 윤현우**

**1. 아래의 선형계획모형을 “DA(2019-02)과제(01).xlsx”파일의 “1번” sheet에 실습하고, 아래 물음에 답하시오.**

|  |
| --- |
| **MIN Z = 2X1 + 1.5X2 + 4X3**  **s.t. 0.5X1 – 3X2 + 3X3 ≥ 54**  **2X1 + 4X2 + 4X3 ≤ 120**  **3X1 + 2X2 + X3 = 90**  **X1, X2, X3 ≥ 0** |

**1) 해답 보고서와 민감도 보고서를 포함하여 해찾기를 수행하시오.**

**2) 최적해는 어떻게 나타나는가? 아래에 답을 typing 하시오.**

1. **X1: 25.41**
2. **X2: 0**
3. **X3: 13.76**

**3) 최적해 적용 시 목적함수는 얼마인가? 아래에 답을 typing 하시오.**

**목적함수: 105.8824**

**2. 매년 한 신문사는 다양한 기준에 근거하여 전국 4년제 대학들에 대한 평가를 실시하고 있다. 그러나 기존의 평가방식은 외형적인 크기에 초점이 맞춰져 있어 규모가 작은 대학들에게 불리하고, 또 일률적으로 부여된 가중치가 각 대학의 특수성을 제대로 반영하지 못한다는 문제가 지적되었다. 이에 당신은 DEA를 이용하여 중소규모 특성화 대학들의 효율성을 평가하고자 한다. 다음은 서울에 소재한 과학기술분야 특성화 대학교 6개 대학에 대한 투입요소 3개와 산출요소 3개에 대한 자료를 조사하여 정리한 것이다.**

* **Input1: 학생 1인당 교수 수(명) – 전체 교수 수 / 재학생 수**
* **Input2: 학생 1인당 예산(백만 원) – 총 예산 / 재학생 수**
* **Input3: 개설 강의 수(과목) – 연간 총 개설 강의 수**
* **Output1: 재학생 만족도(%)**
* **Output2: 대기업 취업률(%)**
* **Output3: 학생 1인당 상장기업 임원 수(명)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **대학** | **Input1** | **Input2** | **Input3** | **Output1** | **Output2** | **Output3** |
| **A** | **0.0533** | **40.45** | **629** | **62** | **45.1** | **0.1711** |
| **B** | **0.0513** | **114.66** | **1,089** | **58** | **39.0** | **0.2578** |
| **C** | **0.0444** | **23.91** | **489** | **71** | **55.1** | **0.1459** |
| **D** | **0.0684** | **32.05** | **536** | **57** | **34.1** | **0.1653** |
| **E** | **0.0258** | **40.47** | **309** | **55** | **19.6** | **0.0423** |
| **F** | **0.0257** | **21.82** | **401** | **65** | **32.8** | **0.0784** |

**1) DEA를 이용하여 DA(2019-02)과제(01).xlsx”파일의 “2번\_A” sheet를 바탕으로 6개 대학의 효율성을 평가하고, 비효율적으로 판명된 대학의 구조조정 방안을 제시하시오.**

**- 풀이: 6개 대학의 효율성을 평가한 결과 비효율적인 대학은 A대학이며 나머지 5개 대학은 효율적이라는 결과가 나왔다. 비효율적으로 판명된 A 대학의 상대적 효율성은 0.9386이며 구조조정 방안은 아래와 같다.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 대학 B | 대학 C | 가상의 효율적 단위 (구조조정 목표) | 비교 | 대학 A | 구조 조정 |
| 투입물 1 | 0.0513 | 0.0444 | 0.05002707 | < | 0.0533 | -0.00327293 |
| 투입물 2 | 114.66 | 23.91 | 33.48466805 | < | 40.45 | -6.965331952 |
| 투입물 3 | 1089 | 489 | 590.3757374 | < | 629 | -38.6242626 |
| 산출물 1 | 58 | 71 | 78.19105639 | > | 62 | 16.19105639 |
| 산출물 2 | 39 | 55.1 | 60.22865632 | > | 45.1 | 15.12865632 |
| 산출물 3 | 0.2578 | 0.1459 | 0.1711 | = | 0.1711 | 0 |

**즉, A 대학의 경우 투입물 1을 0.0033단위, 투입물 2를 6.9653단위, 투입물 3을 38.6243단위 감소시키고, 산출물 1은 16.1911단위, 산출물 2는 15.1287단위 증가시킴으로써 효율적 단위를 변모할 수 있다. (산출물 3에 대한 구조조정 목표와 대학 A의 기존 단위 간 오차는 아주 작은 수 6.63358E-15 이므로 0으로 간주)**

**2) 당신의 보고를 방안을 검토하던 상사는 “[학생 1인당 상장기업 임원 수]를 포함시키는 것이 타당할까? 대학 중 역사가 짧은 대학의 경우 졸업생의 나이가 상장기업 임원의 통상적인 나이에 아직은 미치지 못할 텐데?”라고 지적 하였다. 이에 당신은 산출물 3번을 제외하고 다시 분석을 실시했다. DA(2019-02)과제(01).xlsx”파일의 “2번(수정)\_A” sheet를 바탕으로 6개 대학의 효율성을 평가하고, 비효율적으로 판명된 대학의 구조조정 방안을 제시하시오.**

**- 풀이: 6개 대학의 효율성을 평가한 결과 비효율적인 대학은 A, B, D 대학이며 나머지 C, E, F 대학은 효율적이라는 결과가 나왔다. 비효율적으로 판명된 A, B, D 대학 각각의 상대적 효율성은 0.6785, 0.5957, 0.6804이며 구조조정 방안은 아래와 같다.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 대학 C | 대학 F | 가상의 효율적 단위 (구조조정 목표) | 비교 | 대학 A | 구조 조정 |
| 투입물 1 | 0.0444 | 0.0257 | 0.036165518 | < | 0.0533 | -0.017134482 |
| 투입물 2 | 23.91 | 21.82 | 21.40275946 | < | 40.45 | -19.04724054 |
| 투입물 3 | 489 | 401 | 426.7938207 | < | 629 | -202.2061793 |
| 산출물 1 | 71 | 65 | 63.60462015 | > | 62 | 1.604620147 |
| 산출물 2 | 55.1 | 32.8 | 45.1 | = | 45.1 | 0 |

**A 대학의 경우 투입물 1을 0.0171단위, 투입물 2를 19.0472단위, 투입물 3을 202.2062단위 감소시키고, 산출물 1은 1.6046단위 증가시킴으로써 효율적 단위를 변모할 수 있다.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 대학 F | 가상의 효율적 단위 (구조조정 목표) | 비교 | 대학 B | 구조 조정 |
| 투입물 1 | 0.0257 | 0.030557927 | < | 0.0513 | -0.020742073 |
| 투입물 2 | 21.82 | 25.9445122 | < | 114.66 | -88.7154878 |
| 투입물 3 | 401 | 476.7987805 | < | 1089 | -612.2012195 |
| 산출물 1 | 65 | 77.28658537 | > | 58 | 19.28658537 |
| 산출물 2 | 32.8 | 39 | = | 39 | 0 |

**B 대학의 경우 투입물 1을 0.0207단위, 투입물 2를 88.7155단위, 투입물 3을 612.2012단위 감소시키고, 산출물 1은 19.2866단위 증가시킴으로써 효율적 단위를 변모할 수 있다. (산출물 2에 대한 구조조정 목표와 대학 B의 기존 단위 간 오차는 아주 작은 수 -8.52651E-14 이므로 0으로 간주)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 대학 C | 대학 E | 대학 F | 가상의 효율적 단위 (구조조정 목표) | 비교 | 대학 D | 구조 조정 |
| 투입물 1 | 0.0444 | 0.0258 | 0.0257 | 0.028377726 | < | 0.0684 | -0.040022274 |
| 투입물 2 | 23.91 | 40.47 | 21.82 | 21.8079558 | < | 32.05 | -10.2420442 |
| 투입물 3 | 489 | 309 | 401 | 364.713395 | < | 536 | -171.286605 |
| 산출물 1 | 71 | 55 | 65 | 57 | = | 57 | 0 |
| 산출물 2 | 55.1 | 19.6 | 32.8 | 34.1 | = | 34.1 | 0 |

**D 대학의 경우 투입물 1을 0.04단위, 투입물 2를 10.242단위, 투입물 3을 171.2866단위 감소시킴으로써 효율적 단위를 변모할 수 있다.**

**3. 아래는 Y고등학교 3학년 학생 중 표본으로 추출된 10명의 2019년 9월 모의고사 시험성적 자료이다. 이를 바탕으로 RStudio 실습을 실시하시오.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ID** | **Math** | **Science** | **English** | **Korean** |
| **1** | **61** | **72** | **82** | **59** |
| **2** | **87** | **59** | **91** | **77** |
| **3** | **92** | **69** | **65** | **72** |
| **4** | **65** | **71** | **80** | **68** |
| **5** | **64** | **53** | **86** | **92** |
| **6** | **88** | **91** | **95** | **73** |
| **7** | **75** | **62** | **71** | **59** |
| **8** | **55** | **76** | **63** | **94** |
| **9** | **79** | **59** | **93** | **75** |
| **10** | **90** | **88** | **52** | **67** |

**1) 위 자료로 데이터 프레임을 생성하고, View() 함수를 이용해 생성된 데이터 프레임을 조회하시오.**

> ID <-c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)

> Math <-c(61, 87, 92, 65, 64, 88, 75, 55, 79, 90)

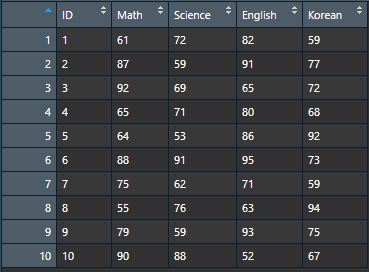
> Science <-c(72, 59, 69, 71, 53, 91, 62, 76, 59, 88)

> English <-c(82, 91, 65, 80, 86, 95, 71, 63, 93, 52)

> Korean <-c(59, 77, 72, 68, 92, 73, 59, 94, 75, 67)

> data\_Assignment <-data.frame(ID, Math, Science, English, Korean)

> View(data\_Assignment)



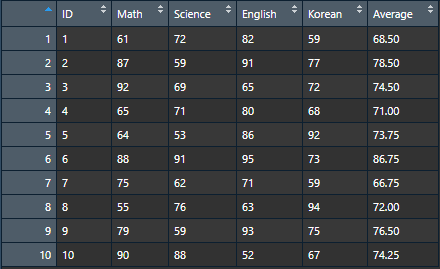
**2) 위 데이터 프레임을 “과제01.csv”파일로 저장하시오.**

> write.csv(data\_Assignment, file = "C:/DA2019/R\_practice/assignment/과제01.csv")

**3) 위 1) 에서 생성된 데이터 프레임에 각 학생의 평균점수로 Average라는 이름의 변수를 추가하시오.**

> data\_Assignment$Average <- (data\_Assignment$Math + data\_Assignment$Science + data\_Assignment$English + data\_Assignment$Korean)/4

> View(data\_Assignment)



**4) describe() 함수를 사용하여 각 과목의 요약통계량을 구하시오.**

> View(data\_Assignment)

> library(psych)

> describe(Math)

vars n mean sd median trimmed mad min max range skew kurtosis se

X1 1 10 75.6 13.57 77 76.12 18.53 55 92 37 -0.15 -1.78 4.29

> describe(Science)

vars n mean sd median trimmed mad min max range skew kurtosis se

X1 1 10 70 12.48 70 69.5 14.08 53 91 38 0.37 -1.28 3.95

> describe(English)

vars n mean sd median trimmed mad min max range skew kurtosis se

X1 1 10 77.8 14.47 81 78.88 16.31 52 95 43 -0.37 -1.41 4.58

> describe(Korean)

vars n mean sd median trimmed mad min max range skew kurtosis se

X1 1 10 73.6 11.89 72.5 72.88 7.41 59 94 35 0.48 -1.09 3.76

**5) “Y고등학교 3학년 학생의 과학성적은 평균 76점이다.”라는 과학교사의 주장을 유의수준 0.05에서 검정하시오. (가설과 함께 분석 결과를 아래에 typing 하시오.)**

**- 가설**

**귀무가설(H0): μ = 76**

**대립가설(H1): μ != 76**

> t.test(Science, mu = 76)

One Sample t-test

data: Science

t = -1.5202, df = 9, p-value = 0.1628

alternative hypothesis: true mean is not equal to 76

95 percent confidence interval:

61.07156 78.92844

sample estimates:

mean of x

70

**- 가설검정결과: 표본 10개의 평균은 70점으로 나타났으며, p-value가 0.1628로 유의수준 0.05보다 크다. 즉, μ = 76이라고 할 수 있다. 결과적으로 H0를 채택한다(= H0을 기각할 수 없다).**

**(하나의 Script 창에 위 실습내용을 작성하고, 작성한 Script를 “과제01.R”로 저장하시오.)**