**Simposio Internacional de Industria y Energía**

**Software for calculating technical efficiency by Data Envelopment Analysis (DEA)**

**DrC. Andrey Vinajera-Zamora 1, MsC. Lissette Concepción-Maure 2,**

**DrC. Norge Coello-Machado3, DrC. Elke Glistau4**

1. Universidad Central “Marta Abreu” de Las Villas, Cuba. [andreyvz@uclv.edu.cu](mailto:andreyvz@uclv.edu.cu)
2. Universidad Central “Marta Abreu” de Las Villas, Cuba. [lissette@uclv.cu](mailto:lissette@uclv.cu)
3. Universidad Central “Marta Abreu” de Las Villas, Cuba. [norgec@uclv.edu.cu](mailto:norgec@uclv.edu.cu)
4. Otto von Guericke University Magdeburg. Germany. [elke.glistau@ovgu.de](mailto:elke.glistau@ovgu.de)

**Abstract**

The technical efficiency is an important element for growing and development of companies, organizations and supply chains. Many researches calculate the technical efficiency for evaluation and diagnosis the different variable using Data Envelopment Analysis (DEA) model. Researchers can use different programs to solve the different DEA model. To use this software, knowledge about their operation is necessary to obtain the model solution and calculate the efficiency; in other cases, they have commercial characteristics that limit access to a few people. On the other hand, non-commercial programs have some limitations according to the quantity of DMU and/or input/outputs. According to this problem, the main objective of this paper is to propose an informatics tool DEA 1.0 (developed with Python 3.6) which one only needs the input data to make all calculation and get technical efficiency. Also, identify the Making Decisions Unit (DMU) with lower technical efficiency and higher one. At the same time, this tool allows the export of the table results in a Microsoft Excel file or as code for LATEX.

**Keywords:** Data Envelopment Analysis; DEA; Software; Optimization’s Tool; Technical Efficiency; MLB Pitching’s Performance

**1. Introduction**

Efficiency measurement has been a subject of tremendous interest as organizations have struggled to improve productivity. Reasons for this focus were best stated fifty years ago by Farrell (1957) in his classic paper on the measurement of productive efficiency. Twenty years after Farrell’s seminal work, and building on those ideas, Charnes et al. (1978), responding to the need for satisfactory procedures to assess the relative efficiencies of multi-input multi-output production units, introduced a powerful methodology which has subsequently been titled data envelopment analysis (DEA) (Cook & Seiford, 2009).

DEA is a linear programming technique for measuring the relative efficiency of decision making units (DMU) on the basis of multiple inputs and multiple outputs. The efficiency is measured in a bounded ratio scale by the fraction “weighted output” to “weighted input”. The inputs and outputs are assumed to be continuous positive variables and the weights are estimated in favor of each evaluated unit so as to maximize its efficiency (Smirlis et al., 2006).

In DEA, the entity (school, organization, hospital, business firm, etc.) under study is called DMU. The definition of DMU is rather lose to allow flexibility in its use over a wide range of possible applications. Generically a DMU is regarded as the entity responsible for converting inputs into outputs and whose performances are to be evaluated (Cooper et al., 2001; Cooper et al., 2006a, 2006b). Performances of DMUs are measured from the optimistic point of view, namely, each DMU seeks a set of weights that is the most favorable to itself to maximize its efficiency (Azizi & Wang, 2013).

DEA have been used to measure efficiency in several researches such as Wu, Liang y Chen (2009), Sadjadi et al. (2011), Liu et al. (2012), Yang et al. (2014), Shafiee et al. (2014), Chen y Du (2015), Atici y Podinovski (2015), Shabani et al. (2015) and other in following fields: sport, health, industry, education, etc. However, to apply this methodology it needs basic knowledge in any software for solve linear programming problem (MATLAB, GAMS, MSES). Also, its necessary spend time to run all models and get the solution. On the other hand, in a literature review was found many software (DEA Solver Pro, Frontier Analyst, OnFront, DEA Excel Solver) that can be very useful for calculate efficiency using any DEA model. However, some programs are commercial and other has some limitations according quantity of DMUs and/or inputs/outputs.

Considering previous aspects, the main objective of these paper is to propose an informatics tool DEA 1.0 (developed with Python 3.6) which only needs the input data to make all calculation and get technical efficiency.

**2. Theoretical background**

The productivity of a particular productive unit is defined as the relationship between the results obtained and the resources involved in its production. The evaluation of productivity is useful when the productive unit to which we refer has the ability to decide to modify well the amount of each one of the resources that are being used, for a quantity of produced results. For this reason, to the productive unit is added the qualifier of decision with the name of Decision Making Unit (DMU).

Modern efficiency measurement begins with Farrell (1957) who drew upon the work of Debreu (1951) and Koopmans (1951) to define a simple measure of firm efficiency which could account for multiple inputs. He proposed that a component inside the efficiency firm is technical efficiency, which reflects the ability of a firm to obtain maximal output from given set of inputs. This is the main principle of DEA which is the non-parametric technique for the performance assessment of a set of Decision Making Units, each of which consumes multiple inputs to produce multiple outputs. The CCR Charnes et al. (1978) and Banker et al. (1984) models are two basic radial models under constant returns to scale (RTS) and variable RTS, respectively.

Actually, the classic DEA models developed with the assumption that all inputs and outputs are non-negative, whereas, this assumption is very restrictive and is not always hold in the real world. For example, the net profit as one of the components of output vector can take a negative value at some DMUs. So, the need to handle negative data and adapt the DEA models has been an attractive issue in DEA literature (Allahyar & Rostamy-Malkhalifeh, 2015).

As the field of Data Envelopment Analysis has grown and blossomed, so have the varieties of models, data, and types of analyses. Similarly, as DEA software technology has emerged from its academic roots into production usage, it has been accompanied by expectations of advanced modeling options and professional implementations, including graphical user interfaces, interoperability with other applications, and the ability to quickly evaluate large populations.

This epigraph surveys the best of the commercial and non-commercial DEA tools available today in table 1 according with . Provided are descriptions of individual packages, comparisons of their features and capabilities, and links to further information on each. This comparison is based in seven variables: available models, key DEA features and capabilities, user interface, reporting, documentation and support, and availability.

Table 1. Comparison of computer tools for DAE analysis

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | COMMERCIAL | | | | NON-COMMERCIAL | | |
|  | DEA Solver Pro 4.0 | Frontier Analyst 3.1.5 | OnFront 20 | Warwick DEA | DEA Excel Solver | DEAP | EMS |
| **MODELS** | | | | | | | |
| NIRS,NDRS,GRS | x |  | x |  | x |  | x |
| Additive/slack-based method | x |  |  | x | x |  | x |
| Malmquist | x | x |  | x | x | x | x |
| Non-convex | x |  |  |  | x |  | x |
| Non-radial | x |  |  | x | x |  |  |
| Free-disposal hull | x |  |  |  | x |  | x |
| Cost efficiency | x |  | x |  | x | x |  |
| Revenue efficiency | x |  | x |  | x |  |  |
| Profit, | x |  |  |  | x |  |  |
| Capacity utilization |  |  | x |  |  |  |  |
| Minimum-efficiency |  |  |  |  | x |  |  |
| Value chain |  |  |  |  | x |  |  |
| **FEATURES** | | | | | | | |
| Window/multi-period analysis | x | x | x |  |  |  | x |
| Super-efficiency scores | x |  |  | x | x |  | x |
| Categorical variables | x |  |  |  |  |  |  |
| Variable priorities |  |  |  | x |  |  |  |
| Sensitivity analysis |  |  |  |  | x |  |  |
| Disposability controls |  |  | x |  | x |  |  |
| Scenario comparison |  |  | x |  |  |  |  |
| **PLATFORM AND INTEROPERABILITY** | | | | | | | |
| Input file types: | SS, manual | TXT,Clip, SS, SS select, SPSS, manual | TXT, Clip, manual | TXT | SS, manual | TXT | TXT, SS |
| Output file types | SS | SS, TXT, PDF, HTML | TXT | TXT | SS | TXT | TXT |
| **USER INTERFACE** | | | | | | | |
| GU I | Excel | x | x | x | Excel |  | x |
| Spreadsheet format | Excel | x | x | x | Excel |  |  |
| individual observation editing | Excel | x | x | x | Excel |  |  |
| Data and results sorting | Excel | x | x | limited capability | Excel |  |  |
| **REPORTING** | | | | | | | |
| Custom reports | Excel | optional |  | limited capability | Excel |  |  |
| Efficiency scores report | x | x | x | x | x | x | x |
| Projected/target factors report | x | x |  | x | x | x |  |
| Optimal factor weights report | x | x | x | x | x | x | x |
| Efficient reference sets report | x | x | x | x | x | xx | x |
| **DOCUMENTATION AND SUPPORT** | | | | | | | |
| Users guide | x | x | x | x | x | x | x |
| Built-in help |  | x | x |  |  |  |  |
| Technical support | x | x |  | x |  |  |  |
| Web site tech info |  | x |  |  | x |  |  |
| **AVAILABILITY** | | | | | | | |
| Free demo available | x | x | x | x |  | x | x |
| Comal license cost | $1600 | $395 | $1750 | $500 |  |  |  |
| Maintenance available |  | x | x | x |  |  |  |

Source: Modified by Barr (2004)

The software packages vary widely in their selection of available DEA models. The classics—CCR/CRS and BCC/VRS—are universally included, but others vary by package. Evaluation Criteria Category Models contains a non-exhaustive list of DEA models and indicates those that are included with each code. There are variants of the models listed in this category —such as input- or output-orientation—listed in the other category as DEA features. All codes allow selection of input or output orientation (and possibly non-oriented).

User interfaces can range from elaborate graphical user interfaces (GUIs) to simple command-line controls. Two of these DEA packages use Microsoft Excel as a GUI for data entry, reporting, and graphics and its Solver tool as a linear program optimizer. The advantages of Excel-based products are: input and output are simplified, reports and graphs can be easily added by the developer and user, interoperability with other Microsoft products is straightforward, and connection to enterprise data sources via ODBC is possible. The disadvantages of the Excel environment are: the default Excel Solver limits application to small populations, larger instances require purchase of a replacement optimizer, and the interpreted execution speed is typically slower than compiled, stand-alone applications.

These software require the user to have knowledge about its operation and programming. As well as a limited number in number of DMU inputs / outputs.

The commercial packages explored are: DEA-Solver-Pro 4.0 from SAITECH, Frontier Analyst®8 3.1.5 from Banxia Software, OnFront 2.02 from Economic Measurement and Quality Corporation, and Warwick DEA from Warwick University. One problem in this area is the non-disclosure of the source code of the designed programs.

The design of the DEA v1.0 software is shown in the following epigraph. DEA was registered in CENDA (national center of copyright in Cuba) referred this free software to the registration 0520-02-2019.

**3. Results and discussion**

The DEA v1.0 software, was written in Python programming environment. It has three languages (English, Spanish and German) interface, doesn't require installation and is registered. The program uses DEA model (CCR and BCC with input and output orientation) and allow calculating efficiency of any quantity of DMU. To show how the program works, we will use an example with three DMUs, two inputs, one output and the efficiency will be calculated by CCR-Input. The welcome window of software after we start the program is figure 1.

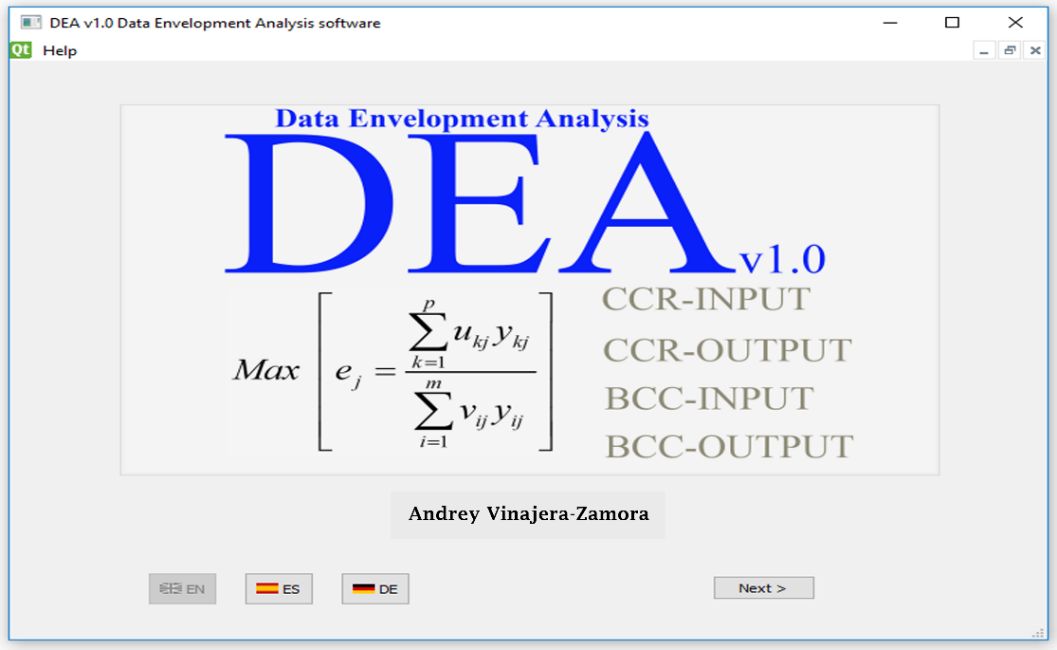


Figure 1: Welcome window of software DEA v1.0

After we press the button Next, we move to the Data entry window (DeW) and enter the data (figure 2).

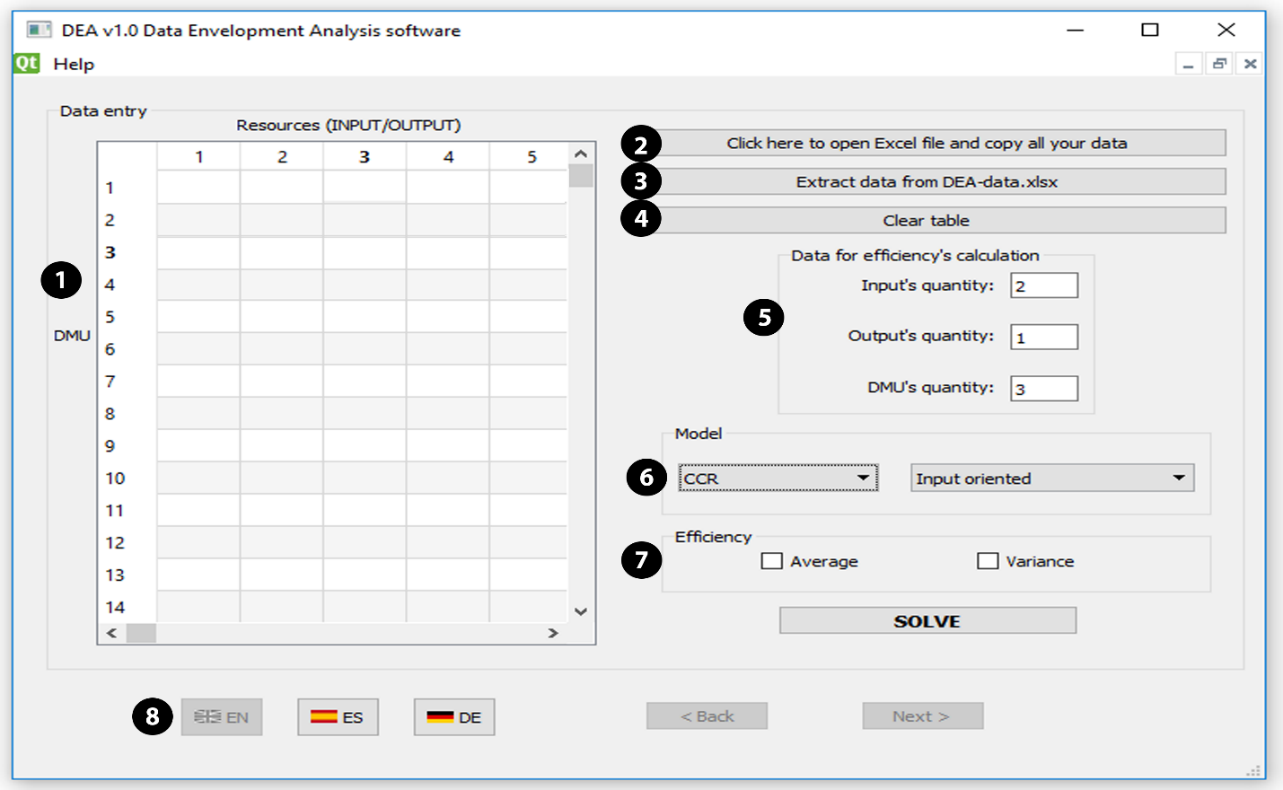


Figure 2: Data entry and configuration window of software DEA v1.0.

1. **Data entry**

To enter data into program both by hand as well as to extract it from Excel file.

1. By hand: Typing cell by cell on table Resources (INPUT/OUTPUT) located in (DeW (1)). The table can be cleared by click on (DeW (4)).
2. Extracted from Microsoft Excel file: For import data you have to follow those steps:

* Open the XLSX file (Microsoft Excel) by clicking on button (DeW (2)) to open DEA-data.XLSX file as show in figure 3. After, you must to paste all your data in this excel file (DEA-data.XLSX) you must save the document.
* Extract data. After save excel file, click on button (DeW (3)) and all your data will appear on table (DeW (1)) in figure 2.
* **Configuring model to be applied**

The dimension of model (quantity of input, output and Decisions Making Units (DMU)) can by specified on text boxes located in (DeW (5)). The kind of model can be selected by (DeW (6)), where the user can choose between CCR or BCC model with input or output orientation. In (DeW (7)), can be checked average or variance to calculate those parameters with efficiency. Initially, the button "**SOLVE**" won’t be activated and will be activated when the data and model configuration is ready. In this time, you will be able to click it and the solution will appear in form shown in **Result** form (Figure 5).

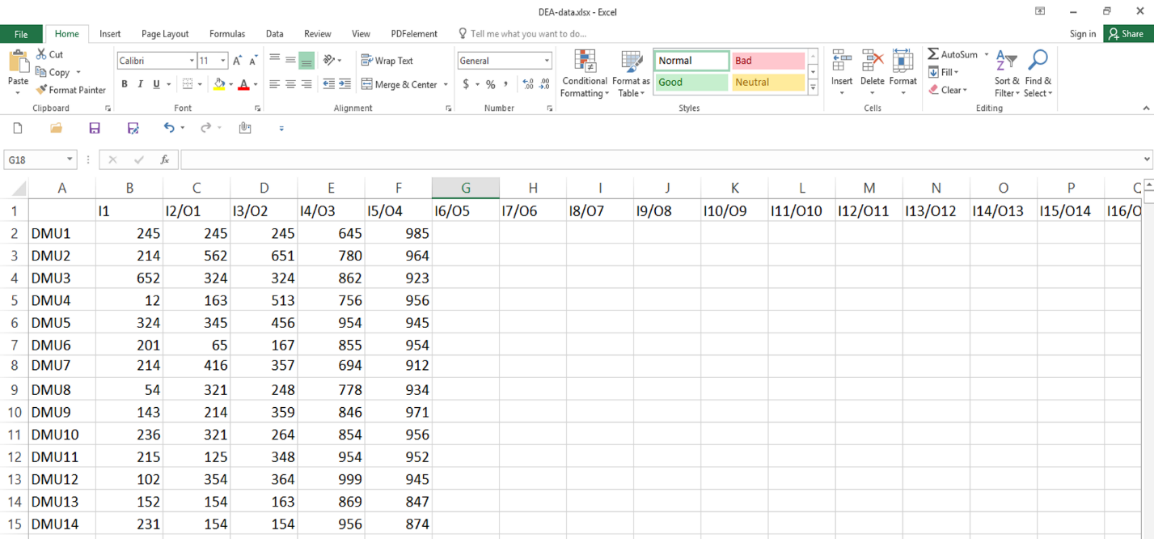


Figure 3: Results exported as excel file

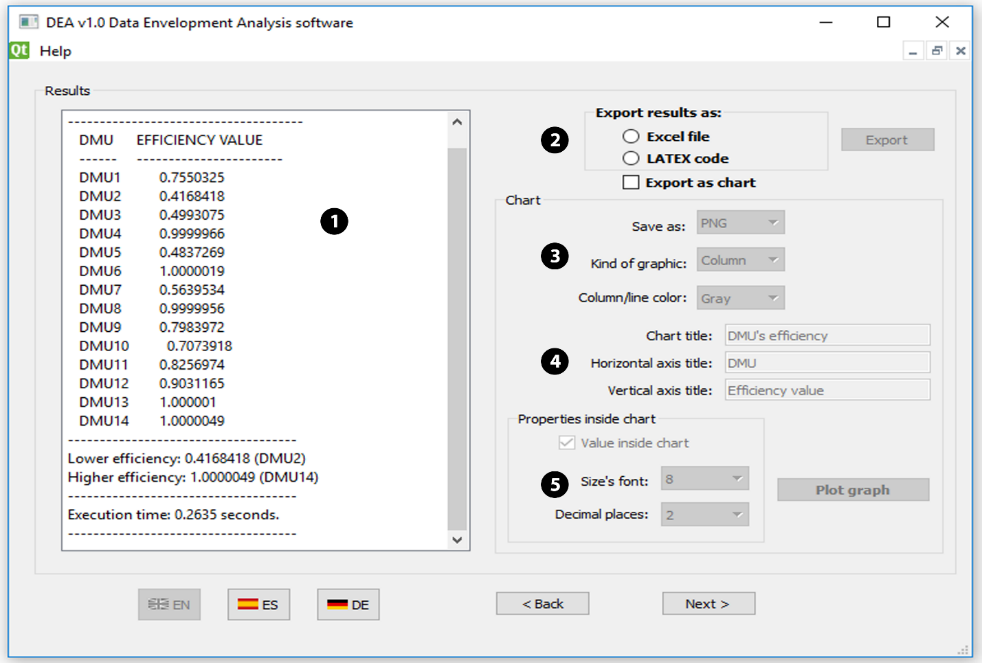
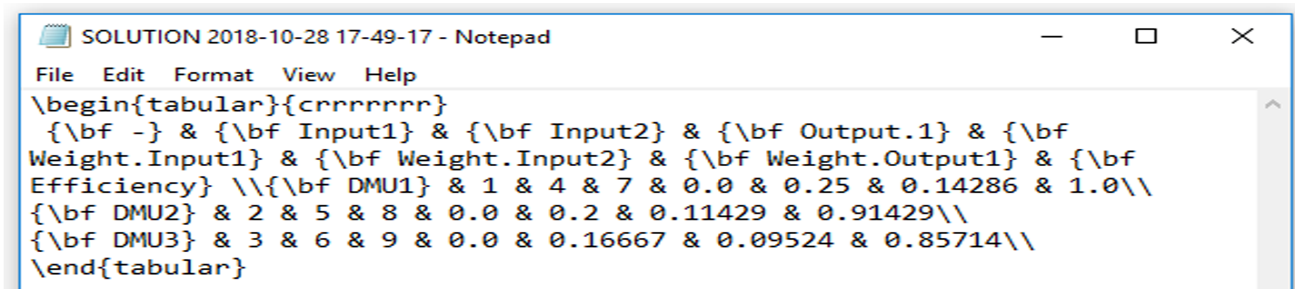
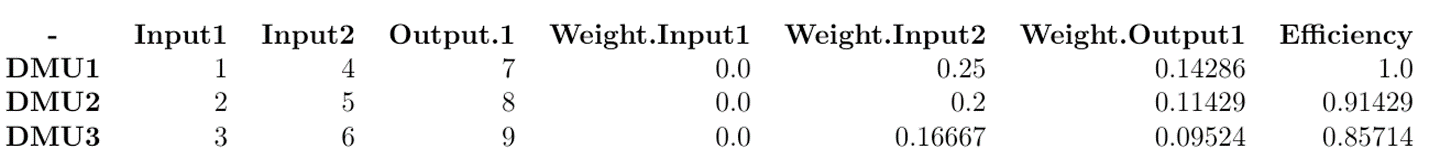
****

Figure 5: Results window of software DEA v1.0

On left side in Results window (RW), there is a panel located in (**RW (1)**) that shows all information (solution, error message, calculation time, etc.). On right side are the options to export results. In **RW (2)** on this form you can export the results as excel file (see figure 6). The result can be exported as LATEX code in text file (figure 7) or as chart in **RW (3)**.



**-A-**



**-B-**

Figure 7: LATEX code (-A-) and table after compiled (-B-) in any software editor of TEX file

In those options, you can select type of \_le (png, jpg, svg, ps, pdf), kind of graphics (line or column) and color of column or line (black, blue, red, yellow, green, brown, gray, cyan and purple). In **RW (4)** will be configured the title of chart, horizontal and vertical axis. And in **RW (5)** can be changed the size fonts inside chart and decimal places of efficiency value inside chart (see figure 8).

**Example**

# For testing the software, we take some data from Major League Baseball corresponding to 2018 MLB Team pitching Statistics located in <https://www.baseball-reference.com/leagues/MLB/2018.shtml>. The pitching efficiency will be considered in this paper as the capacity of the team’s pitchers to don’t earn runs in tied games or the capacity to preserve any advantage when team is winning. According to this definition was selected for DEA model as input: Bases on Balls (BB), homeruns (HR), Hit by a Pitch (HBP) and Fielding Independent Pitching (FIP). The selected outputs were Wins (W) and Saves (SV) as show in table 2. The data corresponding to inputs and outputs are in green cells.

Table 2. Statistics data for pitching by team

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **BB** | **HR** | **HBP** | **W** | **SV** | **Efficiency** | **ERA** | **POS by Wins** |
| **ARI** | 522 | 174 | 57 | 82 | 39 | **0.832** | 3.72 | 14 |
| **ATL** | 635 | 153 | 52 | 90 | 40 | **1.000** | 3.75 | 11 |
| **BAL** | 589 | 234 | 67 | 47 | 28 | **0.437** | 5.18 | 30 |
| **BOS** | 512 | 176 | 84 | 108 | 46 | **0.906** | 3.75 | 1 |
| **CHC** | 622 | 157 | 66 | 95 | 46 | **0.953** | 3.65 | 6 |
| **CHW** | 653 | 196 | 89 | 62 | 34 | **0.552** | 4.84 | 28 |
| **CIN** | 532 | 228 | 50 | 67 | 38 | **0.775** | 4.63 | 24 |
| **CLE** | 407 | 200 | 59 | 91 | 41 | **0.969** | 3.77 | 8 |
| **COL** | 525 | 184 | 52 | 91 | 51 | **1.000** | 4.33 | 9 |
| **DET** | 491 | 216 | 60 | 64 | 37 | **0.676** | 4.58 | 26 |
| **HOU** | 435 | 152 | 66 | 103 | 46 | **1.000** | 3.11 | 2 |
| **KCR** | 549 | 205 | 51 | 58 | 33 | **0.660** | 4.94 | 29 |
| **LAA** | 546 | 205 | 65 | 80 | 35 | **0.706** | 4.15 | 18 |
| **LAD** | 422 | 179 | 61 | 92 | 48 | **0.998** | 3.38 | 7 |
| **MIA** | 605 | 192 | 71 | 63 | 30 | **0.538** | 4.76 | 27 |
| **MIL** | 553 | 173 | 61 | 96 | 49 | **0.949** | 3.73 | 5 |
| **MIN** | 573 | 198 | 80 | 78 | 37 | **0.614** | 4.5 | 19 |
| **NYM** | 484 | 185 | 71 | 77 | 41 | **0.728** | 4.07 | 20 |
| **NYY** | 494 | 177 | 58 | 100 | 49 | **1.000** | 3.78 | 3 |
| **OAK** | 474 | 184 | 53 | 97 | 44 | **1.000** | 3.81 | 4 |
| **PHI** | 500 | 171 | 60 | 80 | 44 | **0.823** | 4.14 | 17 |
| **PIT** | 497 | 174 | 69 | 82 | 40 | **0.753** | 4 | 15 |
| **SDP** | 519 | 185 | 68 | 66 | 36 | **0.626** | 4.4 | 25 |
| **SEA** | 400 | 195 | 72 | 89 | 60 | **1.000** | 4.13 | 12 |
| **SFG** | 524 | 156 | 45 | 73 | 36 | **0.903** | 3.95 | 21 |
| **STL** | 593 | 144 | 67 | 88 | 43 | **0.970** | 3.85 | 13 |
| **TBR** | 501 | 164 | 53 | 90 | 52 | **1.000** | 3.74 | 10 |
| **TEX** | 491 | 222 | 72 | 67 | 42 | **0.667** | 4.92 | 23 |
| **TOR** | 551 | 208 | 67 | 73 | 39 | **0.665** | 4.85 | 22 |
| **WSN** | 487 | 198 | 76 | 82 | 40 | **0.722** | 4.04 | 16 |

After run the software (with three inputs, two outputs, thirty DMUs and CCR model with input orientation) the efficiency by team can be seen in figure 8. The worst teams was Baltimore (0.43), Chicago (0.55) and Miami (0.53). At same time, in table 2, orange cell represents the lowest efficiency, blue cells are the team with worst ERA and gray cell represents the worst teams according their wins. As we can see, the teams with lowest efficiency, worst ERA and worst wins corresponding to the same teams.

Figure 8. Efficiency by team

**4. Conclusion**

The article presents the latest version of the software DEA. Software is dedicated for calculate technical efficiency by Data Envelopment Analysis model. The use of software makes easier and faster the calculating. The software is being wildly tested. Actually, the author is working to calculate efficiency by dual model using those DEA models. Also, for future versions the software will have options to save and open files previously saved and configured.

**5. Bibliography**

Allahyar, M., & Rostamy-Malkhalifeh, M. (2015). Negative data in data envelopment analysis: Efficiency analysis and estimating returns to scale. *Computers & Industrial Engineering*, *82*, 78–81. https://doi.org/10.1016/j.cie.2015.01.022

Atici, K. B., & Podinovski, V. V. (2015). Using data envelopment analysis for the assessment of technical efficiency of units with different specialisations: An application to agriculture. *Omega*, *54*, 72–83. https://doi.org/10.1016/j.omega.2015.01.015

Azizi, H., & Wang, Y.-M. (2013). Improved DEA models for measuring interval efficiencies of decision-making units. *Measurement*, *46*(3), 1325–1332.

Banker, R., Charnes, A., & Cooper, W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, *30*(9), 1078–1092. Retrieved from http://pubsonline.informs.org/doi/abs/10.1287/mnsc.30.9.1078

Barr, R. S. (2004). Dea Software Tools and Technology BT - Handbook on Data Envelopment Analysis. In W. W. Cooper, L. M. Seiford, & J. Zhu (Eds.) (pp. 539–566). Boston, MA: Springer US. https://doi.org/10.1007/1-4020-7798-X\_18

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444.

Chen, Y., & Du, J. (2015). Super-Efficiency in Data Envelopment Analysis. In *Data Envelopment Analysis. A Handbook of Models and Methods* (pp. 381–414). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4899-7553-9\_14

Cook, W. D., & Seiford, L. M. (2009). Data envelopment analysis (DEA)–Thirty years on. *European Journal of Operational Research*, *192*(1), 1–17.

Cooper, W. W., Li, S., Seiford, L. M., Tone, K., Thrall, R. M., & Zhu, J. (2001). Sensitivity and stability analysis in DEA: some recent developments. *Journal of Productivity Analysis*, *15*(3), 217–246.

Cooper, W. W., Seiford, L. M., & Tone, K. (2006a). *Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software*. Springer.

Cooper, W. W., Seiford, L. M., & Tone, K. (2006b). The basic CCR model. In *Introduction to Data Envelopment Analysis and Its Uses: With DEA-Solver Software and References* (pp. 21–39).

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica: Journal of the Econometric Society*, *1951*, 273–292.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 253–290.

Koopmans, T. C. (1951). An analysis of Production as an Efficient Combination of Activities. In Wiley (Ed.), *Activity analysis of Production and Allocation* (T.C. Koopm). New York.

Liu, J. S., Lu, L. Y. Y., Lu, W.-M., & Lin, B. J. Y. (2013). Data envelopment analysis 1978–2010: A citation-based literature survey. *Omega*, *41*(2013), 3–15. https://doi.org/10.1016/j.omega.2010.12.006

Sadjadi, S. J., Omrani, H., Abdollahzadeh, S., Alinaghian, M., & Mohammadi, H. (2011). A robust super-efficiency data envelopment analysis model for ranking of provincial gas companies in Iran. *Expert Systems with Applications*, *38*(9), 10875–10881. https://doi.org/10.1016/j.eswa.2011.02.120

Shabani, A., Torabipour, S. M. R., Farzipoor Saen, R., & Khodakarami, M. (2015). Distinctive data envelopment analysis model for evaluating global environment performance. *Applied Mathematical Modelling*, *39*(15), 4385–4404. https://doi.org/10.1016/j.apm.2014.12.053

Shafiee, M., Hosseinzadeh Lotfi, F., & Saleh, H. (2014). Supply chain performance evaluation with data envelopment analysis and balanced scorecard approach. *Applied Mathematical Modelling*, *38*(21–22), 5092–5112. https://doi.org/10.1016/j.apm.2014.03.023

Smirlis, Y. G., Maragos, E. K., & Despotis, D. K. (2006). Data envelopment analysis with missing values: an interval DEA approach. *Applied Mathematics and Computation*, *177*(1), 1–10.

Wu, J., Liang, L., & Chen, Y. (2009). DEA game cross-efficiency approach to Olympic rankings. *Omega*, *37*(4), 909–918. https://doi.org/http://dx.doi.org/10.1016/j.omega.2008.07.001

Yang, M., Li, Y., Chen, Y., & Liang, L. (2014). An equilibrium efficiency frontier data envelopment analysis approach for evaluating decision-making units with fixed-sum outputs. *European Journal of Operational Research*, *239*(2), 479–489. https://doi.org/10.1016/j.ejor.2014.05.013