Journal of Mathematics and Statistics 1 (2): 103-105, 2005
ISSN 1549-3644

On Selecting "r" Items From "m" Independent Groups

[1]Ezz H. Abdelfattah, [2]Mohamed Hafez and [1]Said F. Tantawy
[1]Faculty of Science, Helwan University, Cairo, Egypt
[2]Faculty of Economic and Political Science, Cairo University, Cairo, Egypt

Abstract: Choosing the best r individuals from m independent groups by the regular way, doesn't consider the competition "between" these groups, but it only considers the competition "within" the groups. In this study we discuss the selection by considering the competition with both between and within the groups, by using a Linear Programming (*LP*) approach and depending on the True Position (*TP*) introduced by Abdelfattah [1], which depends on rescaling the data for each group.

Key words: Linear Programming Approach, True Position

INTRODUCTION

Assume that we have a data consisting of *m* group, each group contains *n* observations (items). Let $\{O_{ij}\}, i = 1,2,...,n; j = 1,2,...,m$, represents the observed data. Assume that it is required to select the best *k=rm* item based on these data, where *r* is an integer less than *n*. The "regular" method of choice is based on ranking the observations "within" each group and then selecting the best *r* item from each group. This way of selection may not ensure the best selection, because we may choose an item and leave a better individual assigned a rank that is less than what it deserve and is better than the selected one. The proposed procedure is based on transforming the observations to its true position (*TP*), by rescaling the data within each group from 1 to *n*, then applying a linear programming approach to select the best items. Here, we will briefly review the *TP* procedure and formulate a *LP* model for two different cases. Also we will apply the two models to the scores gained by the Football teams attending the World Cup that was hold at France in 1998.

A True Position: Let $x_{(1)} < x_{(2)} < ... < x_{(n)}$ represents the order statistics (*OS*) of $x_1, x_2, ... x_n$. Then the *TP* procedure goes as follows [1]:
Assign the rank 1 for the smallest value $x_{(1)}$ and the rank n for the largest value $x_{(n)}$ and let:

$$D_x = x_{(n)} - x_{(1)} \quad (2.1)$$

be the range of the n values and

$$D_R = n - 1 \quad (2.2)$$

be the range of the n ranks. Hence the (ascending) *TP* for $x_{(i)}$ (or x_i); denoted by $TP_a(x_i)$ can be obtained through the equation:

$$TP_a(x_i) = \frac{D_R}{D_x}(x_i - x_{(1)}) + 1 \quad (2.3)$$

Note that the formula (2.3) requires knowing $x_{(1)}$ and $x_{(n)}$ only to obtain $TP_a(x_i)$ for i=*1,2,...,n*. Also, through (2.3), we can obtain the position for each value independently of the position of all the other values. This will avoid comparing each value with 'all' other values to be able to assign the required position, as usual in the ordinary ranking (*OR*) procedure. It is clear that (2.4) assigns the position 1 for $x_{(1)}$ and the position n for $x_{(n)}$, while the ascending position corresponding to $x_{(i)}, 1 < i < n$ will be assigned according to its position relative to both $x_{(1)}$ and $x_{(n)}$.
We use equation (2.3) when ranking the values in ascending order, while ranking the values in descending order requires using the equation:

$$TP_d(x_i) = \frac{D_R}{D_x}(x_{(n)} - x_i) + 1 \quad (2.4)$$

It is clear that (2.4) assigns the position 1 for $x_{(n)}$ and the position n for $x_{(1)}$ while the position corresponding to $x_{(i)}, 1 < i < n$ will be assigned according to its position relative to both $x_{(1)}$ and $x_{(n)}$.
Note that the *TP* procedure - in case of **tied position-** will assign "same" position for same values. The following propositions for the relations between *TP* with the original Observations X and *OR* will be given without proof and the proof can be found at Abdelfattah [1].

Proposition 2.1: *TP* reflect the actual position for X.

Proposition 2.2: *TP* is equivalent to the original data X iff $x_{(1)} = 1$ and $x_{(n)} = n$.

Proposition 2.3:
i) The correlation between X and TP_a is +1.00.
ii) The correlation between X and TP_d is -1.00.

Proposition 2.4: *TP* is equivalent to *OR* iff the observations constitute an arithmetic sequence.

We may look for the nonnumeric data (such as "good, very good, excellent") as if i t is 'equidistant' from each other; and consequently in this case we can use the *TP* as an *OR*.
It is clear also that the relation between the *TP* and the *OR* will be the same as the relation between the original observations X and the *OR*.

Linear Programming Models: According to the criterion of selection, we have two models:

Model 1: Let the number of items to select equals *k*, verifying *k*≤*n m*, where *m* is the number of groups and *n* is the number of items in each group (which is supposed to be the same for each group). This type assumes that the chosen number of items is multiple of the number of the groups at most.
Consider $TR_d(o_{ij})$, the descending true position score of the item o_{ij} as defined by (2.4), as the criterion of selection. We may formulate the following LP model [2] to operate the selection of *r* (≤*n m*) best items. Naturally, the best *r* items are those having least scores (due to descending ranking) and consequently they have minimum sum of scores, so the objective function of the LP model will be defined as :

$$Min \sum_{i,j} TR_d(O_{ij}).X_{ij} \qquad (3.1)$$

This objective function will be constrained by:

$$\sum_{i,j} X_{ij} = k$$
$$X_{ij} \le 1 \quad ,1 \le i \le m;\ 1 \le j \le n \qquad (3.2)$$
$$X_{ij} \ge 0$$

The solution of this LP model insures that the decision variable X_{ij} will take the values 0 or 1 only and mainly,

$$X_{ij} = 1 \quad \textit{if item ij is selected}$$
$$= 0 \quad \textit{otherwise}$$
$$1 \le i \le m\ ,\ 1 \le j \le n.$$

Example 3.1: Consider the following scores given in Table 1, that represents the scores (points) O_{ij} gained after the competition games between 32 teams classified into 8 groups, each of size 4. These scores are the actual scores for the games during the World Cup that hold at France in 1998. (http://www.france98.com/english/competition/groupstandings.htm)

Table 1: Score of Item O_{ij}

Team/Group	1	2	3	4
G1	6	5	4	1
G2	7	3	2	2
G3	9	4	2	1
G4	6	5	4	1
G5	5	5	3	1
G6	7	7	3	0
G7	7	6	3	1
G8	9	6	3	0

Table 2: The Descending True Ranking $TR_d(o_{ij})$

Team/Group	1	2	3	4
G1	1	1.6	2.2	4
G2	1	3.4	4	4
G3	1	2.88	3.63	4
G4	1	1.6	2.2	4
G5	1	1	2.5	4
G6	1	1	2.71	4
G7	1	1.5	3	4
G8	1	2	3	4

The descending true ranking computed for each item in each group is given in Table 2.
Assume that it is required to selected the top 16 teams (*k*=16). The regular (classical) way, was to select the first and the second teams in each of the 8 groups. According to our procedure and applying the LP model (3.1 and 3.2), we obtain the decision variables X_{ij} - by running an OR software [3] which assigns the value 1 for the selected team and 0 for the team that was not selected. The output is shown in Table 3.
This means that the 3rd team of group 1 (Morocco) and the 3rd team of group 4 (Spain) should be chosen instead of the team in the 2nd place of group 2 (Chile) and the team in the 2nd place of group3 (South Africa).
It is necessary to mention that items having X_{ij}=1 in Table 3 are picked by the model due to the order given in Table 3.
The order of selection of items given in Table 4 depends mainly on the priority of inscription of groups. This priority has no effect on the result of selection if TR_d of the last item chosen does not equal any other TR_d of non-chosen items. In case of equality, we have to return to the original data O_{ij} and then choose the

Table 3: The Values of X_{ij}

Team/Group	1	2	3	4
G1	1	1	1	0
G2	1	0	0	0
G3	1	0	0	0
G4	1	1	1	0
G5	1	1	0	0
G6	1	1	0	0
G7	1	1	0	0
G8	1	1	0	0

Table 4: The Order of Selection of Case1

Team/Group	1	2	3	4
G1	**1**	**12**	**15**	
G2	**2**			
G3	**3**			
G4	**4**	**13**	**16**	
G5	**5**	**9**		
G6	**6**	**10**		
G7	**7**	**11**		
G8	**8**	**14**		

item having larger score. In a mathematical formulation, this process could be given as follows:

Define:

$$A = \{TR_d(O_{ij}) \mid X_{ij} = 1\}$$
$$B = \{TR_d(O_{ij}) \mid X_{ij} = 0\}$$

Let:

$$TR_d(O_{ls}) = \max A$$
$$TR_d(O_{rv}) = \max B$$

In case of $TR_d(O_{ls}) = TR_d(O_{rv})$ and if $O_{ls} < O_{rv}$, then put $X_{ls} = 0$ and $X_{rv} = 1$.In other words, we select item *rv* and neglect item *ls*.
Here, we have to mention that -in case of small *m* and *n*- the choice process may be calculated manually by the previous procedure, rather than using the software, otherwise it will be loosing time.

Model 2: If we consider $TR_d(O_{ij})/O_{ij}$ (that is to divide the TR for each cell by the corresponding sore), as another criterion of selection of items rather than $TR_d(O_{ij})$, this will result another order of selection. In this case, we define the LP model as follows:
Define $w_{ij} = TR_d(O_{ij})/O_{ij}$, so that the LP model will be given as:

$$Min \sum_{i,j} w_{ij}.X_{ij} \qquad (3.3)$$

Subject to:

$$\sum_{i,j} X_{ij} = k$$
$$X_{ij} \le 1 \quad ,1 \le i \le m;\ 1 \le j \le n$$
$$X_{ij} \ge 0$$

Applying the LP model (3.3) to the previous example, we obtain the same values of the decision variables X_{ij} as given in Table 3, but with another order of selection. The items (having X_{ij}=1 in Table 3) are picked by the LP model (3.3) due to the order given in Table 5.

Table 5: The Order of Selection for Case 2

Team/Group	1	2	3	4
G1	**7**	**12**	**15**	26
G2	**3**	22	24	24
G3	**1**	17	23	26
G4	**7**	**12**	**15**	26
G5	**9**	**9**	18	26
G6	**3**	**3**	19	too big
G7	**3**	**11**	20	26
G8	**1**	**14**	20	too big

Table 6:

Team/Group	1	2	3	4
G1	**7**	**11**	**15**	26
G2	**3**	18	23	23
G3	**1**	**15**	23	26
G4	**7**	**11**	**15**	26
G5	**11**	**11**	18	26
G6	**3**	**3**	18	31
G7	**3**	**7**	18	26
G8	**1**	**7**	18	31

Note: We may clarify here that if we think about ranking each item of the data given in Table 1 with respect to each other (as if they were in a raw), we will obtain the rank given in Table 6 which is clearly different than the rank assigned by our procedure.

REFERENCES

1. Abdelfattah, Ezz H., 2003. A True Position as A New Way for Monitoring Data. The 15th Annual Conference on Statistics and Computer Modeling in Human and Social Sciences, Faculty of Economics and Political Science, Cairo University, Egypt.
2. Thie, P., 1988. An Introduction to Linear Programming and Game Theory. 2nd Edition, John Wiely & Sons, Inc, NY.
3. Kalvelogen, E. and H.Ti, 1990. Operation Research and Statistics. Urije Universiteit, Amsterdam.