

# Using a K-Means Clustering Algorithm to Examine Patterns of Vehicle Crashes in Before-After Analysis

Raffaele Mauro<sup>1</sup>, Mario De Luca<sup>2</sup> & Gianluca Dell'Acqua<sup>2</sup>

<sup>1</sup> University of Trento, Via Mesiano 7, Trento, Italy

<sup>2</sup> University of Napoli Federico II, Via Claudio 21, Napoli, Italy

Correspondence: Mario De Luca, University of Napoli Federico II, Via Claudio 21, Napoli 80125, Italy. Tel: 39-81-768-3936. E-mail: mario.deluca@unina.it

Received: August 2, 2013

Accepted: September 4, 2013

Online Published: September 11, 2013

doi:10.5539/mas.v7n10p11

URL: <http://dx.doi.org/10.5539/mas.v7n10p11>

## Abstract

The study aims to develop a support procedure to estimate the efficacy of infrastructural interventions to improve road safety. The study was carried out on a 110 km stretch of the A3 highway, in southern Italy. Data from a huge sample concerning traffic, geometry and accidents for two periods of the same duration were compared, for which cluster analysis, and in particular, the “hard c means” binary partition algorithm was employed. Using cluster analysis, all the accidents with strong similarities were aggregated. Then for each cluster, the “cluster representative” accident was identified, to find the average among the various characteristics (geometrical, environmental, accident-related). A “hazard index” was also created for each cluster, whereby it was possible to establish the danger level for each “cluster”. Using this information, an accident prediction model using a multi-variate analysis was produced. This model was used as a support for decision-making on infrastructures and to simulate situations to which the *Before-After* technique could be applied.

**Keywords:** planning, road safety, cluster analysis, after-before analysis

## 1. Introduction and Previous Studies

All policies that affect travel patterns also affect the numbers killed and injured in transport accidents, and conversely, changing the travel patterns may in itself be a way of reducing these numbers. Investigation of this interaction between travel patterns (De Luca & Dell'Acqua, 2012a) and the number of deaths and injuries in transport accidents can benefit greatly from various kinds of data that are already commonly collected in travel surveys (Dell'Acqua, 2012).

A recent study (Ghosh & Savolainen, 2012) examined the factors that affect the time required by the Michigan Department of Transportation Freeway Courtesy Patrol to clear incidents occurring on the southeastern Michigan freeway network. These models were developed using traffic flow data, roadway geometry information, and an extensive incident inventory database. Lord et al. (2010) provided a detailed review of the key issues associated with crash-frequency data as well as the strengths and weaknesses of the various methodological approaches that researchers have used to address these problems. Depaire et al. (2008), always in the field of road safety, demonstrated cluster analysis in order to identify homogeneous classes of accidents that allowed for a very effective analysis. Similarly, KwokSuen et al. (2002) used cluster analysis to group homogeneous data in an experimental analysis to develop an algorithm to estimate the number of road accidents and to assess the risk of accidents. Finally Dell'Acqua et al. (2011a) applied cluster analysis to develop a Decision Support System (DSS) useful to identify the Accident Modification Factors (AMF).

## 2. Methodology and Objectives

The methodology used in this work is based on two techniques:

- a) Cluster Analysis (Hard c-Means Method -H c m);
- b) The “Before-After” approach;

Using “Cluster Analysis” (by means of the special binary partition algorithm “hard c-means”) all accidents with a high degree of resemblance or strong similarities were aggregated (clustered). Then the “cluster representative” accident for each cluster was determined to find the average of all the different characteristics (geometric,

environmental, traffic-related). Then, for each cluster, a *hazard index* was created whereby it was possible to determine the danger rate for each *cluster* in terms of accidents. Using this information (the average characteristics of the “cluster representative” as independent variables, and the “hazard index” as the dependent variable) it was possible to construct, by means of multiple-regression multivariate analysis, an accident prediction model. This model was used to support the choice of infrastructure projects and to simulate situations in which the *Before-After* technique could be applied. In particular, the model was used to simulate “after” situations, with and without any intervention. A comparison of these two situations, simulated by the model, made it possible to assess the effectiveness of interventions in terms of road safety. Detailed descriptions of the two techniques mentioned in this section follow. The study aims to develop a support procedure to estimate the efficacy of infrastructural interventions to improve road safety. The procedure was developed using the two techniques described above in an experiment on a stretch of highway of approximately 110 km.

### 2.1 Cluster Analysis with Hard c-Means Algorithm (H c m)

The principles of this technique are as follows (Tryon, 1939). The aim of the group analysis consists in identifying a specific U partition, in c groups ( $2 \leq c \leq n$ ) of the U collection space constituted by n-elements. The hypothesis upon which this method is based is the following: the elements of the X space, that belong to a group, are characterized by a mathematical affinity and this affinity is greater than the elements of the different groups. Each element in the sample can be schematized as a point identified by m-coordinates, and each coordinate constitutes an attribute of the same element. One of the simpler measures of affinity is represented by the distance measured between two points which belong to the data-space (De Luca et al., 2012b)

### 2.2 Before-After Analysis

In this study, the *Naïve “Four steps/before-after”* study (Hauer, 1997) was used. To overcome the limitations of the “naïve” techniques, the data were subject not only to cluster analysis, but they were also given an accident rate index, whose details are illustrated below, to take into consideration the traffic variable. In this way, it was possible to compare, in the two different comparison configurations (“after” without intervention” and “after”, with intervention), the traffic conditions, the meteorological conditions, the number of vehicles involved, and other quantities, the details of which are given in later sections.

#### 2.2.1 Naïve “Four Steps/Before-After” Study

Let us suppose that an intervention was carried out on 1, 2, ..., j, ..., n. Let us also suppose that in the period preceding the intervention (Before), K(1), K(2), ..., K(j), ..., K(n) accidents occurred; in the “after” period then, the number of accidents was L(1), L(2), ..., L(j), ..., L(n). Since the duration of the periods “before” and “after” (De Luca et al., 2011) may be different from one entity to another (where entity means an intersection, a curve exiting a tunnel with a wet surface, a road/rail crossing with level-crossing, etc), we define the “duration ratio” for entity j as:

$$r_d(j) = \frac{\text{duration of the after period for entity } j}{\text{duration of the before period for entity } j} \quad (1)$$

Let us also define as follows:

$\pi$ , expected number of accidents in the “after” period with no intervention

$\lambda$ , expected number of accidents in the “after” period following a specific intervention.

Specifically:

$$\lambda = \sum L(j) \quad (2)$$

$$\pi = \sum r_d^2(j) \cdot \hat{k}(j) \quad (3)$$

$$VAR\{\lambda\} = \sum L(j) \quad (4)$$

$$\widehat{VAR}\{\pi\} = \sum r_d^2(j) \cdot VAR\{k(j)\} \quad (5)$$

To establish the validity of the effectiveness of the intervention, the indices (6) and (7) are used:

$$\theta = \frac{\frac{\lambda}{\pi}}{1 + \frac{VAR\{\pi\}}{\pi^2}} \quad (6)$$

Defined as the relationship between the degree of safety with the modernization works and what it would have been without them:

$$\delta = \pi - \lambda \quad (7)$$

Defined as the number of accidents expected (or expected accident rate) “after” the intervention.

Let us also define the variation of  $\theta$  and  $\delta$  as:

$$VAR\{\theta\} = \theta^2 \cdot \frac{\frac{VAR\{\hat{\lambda}\}}{\lambda^2} + \frac{VAR\{\hat{\pi}\}}{\pi^2}}{\left[1 + \frac{VAR\{\hat{\pi}\}}{\pi^2}\right]^2} \quad (8)$$

$$\widehat{VAR}\{\delta\} = \widehat{VAR}\{\pi\} + \widehat{VAR}\{\lambda\} \quad (9)$$

In particular, the first index  $\theta$ , takes on the following important meaning:

If  $\theta > 1$ , the intervention has a negative influence on safety

If  $\theta < 1$ , the intervention has no negative influence on safety

### 3. Data Collection

The Stretch analyzed (see Figure 1) belongs to the A3 (Salerno-Reggio Calabria freeway) situated in the south of Italy between distance “170.000 km”, and distance “259.000 km” (Dell’Acqua et al., 2011c).

#### 3.1 Description of the Stretch “Before” the Intervention

The stretch t analyzed (from 30 September 1998 to 30 September 1999) has a slope of between 0% and 4%. The road surface is dense asphalt-flexible type. Geometric data were obtained from map sources to a scale of 1:2000 and 1:10000. The accidents used (totaling about 520 in the whole segment) were made available by the police authorities; traffic data were taken from the archives of the local government administration. The collected data were organized as indicated in Table 1.



Figure 1. Stretch analyzed

### 3.2 Description and Comparison of the Sub-Stretch Where the Infrastructure Operations Took Place

Between the years 1999 and 2001, modernization works were carried out on the stretch (from 218.000 km to 259.000 km).

In the period “after” (from 30 September 2002 to 31 August 2003) there were 85 accidents with 38 injuries and no deaths. On the same stretch in the “Before” period (from 30 September 1998 to 31 August 1999) there were 104 accidents with 58 injuries and 1 fatality.

Table 1. Collecting and organizing data

Distance [Km]	Date	hour	Direction	Slope [%]	Geometric element	Curve Radius [m]	Surface state	Light condition	n. Injured and deaths	Type accident	Location	Driver	ADT [veh./day]
180.1	7/8/99	8:30	S	2	Tangent		dry	Day	0	Skid	Tunnel	Residents	15000
180.75	6/19/99	15:05	N	2	Tangent		dry	Day	0	Skid	Bridge	Residents	15000
181.29	4/25/99	8:00	S	2	Curve	700	dry	Day	2inj.	Skid	Tunnel	Residents	15000

## 4. Data Analysis

### 4.1 Cluster Analysis (H c m Algorithm)

Hard c means algorithm was applied to the variables showed in table 2 (to accidents included in the period from 30 September 1998 to 30 September 1999 and distance from 170.000 to 281.000 km). Table 3 shows the results obtained.

Table 2. Variables introduced in the cluster analysis

Variable	Description	Range	Label	Type of variable
Curving	Inverse of radius (1/R)	[0.00 ÷ 0.0028]	CV	Continuous
Slope	+ Upgrade	[-2.5 ÷ 2.5]	LG	Continuous
	- Downgrade			
Freeway exit	-Absence	1	EXF	Discrete
	-Presence of “freeway exit” with acceleration and deceleration lane planned correctly	1.25		
	-Presence of “freeway exit” with acceleration and deceleration lane not planned correctly	1.5		
Light	Day	1	LC	Discrete
	Night	2		
Paving	Dry	1	PV	Discrete
	Wet	2		

The clusters identified using the “H c m” technique can be considered as a “hazardous zone”. If we analyze the first line of Table 3 (obtained by calculating the mean value within the cluster) we can observe that it has the following average characteristics:

- Variable 1 CV: Low curvature
- Variable 2 LG: Low longitudinal slope
- Variable 3 EXF: Presence of a freeway exit
- Variable 4 LC: Nocturnal light conditions
- Variable 5 PV: Wet road paving

The same can be said of other “clusters” shown in Table 3.

Table 3. Results of cluster analysis

Cluster	Var 1- Curving [1/m]	Var 2- Slope [%]	Var 3 -Freeway exit	Var 4- Light	Var 5- paving	severity	Number of vehicles	Length of the “hazardous Zone”[km]	ADT After without work. (from Sep.2002 to August 2003) [Veh./day]	“Id index” calculated with Eq. (10)
a	0.000272	3	1.5	2	2	1.14	27	1.8	14920	1879
b	0.002547	2.4	1	1.85	1.86	1.27	51	3.6	14840	1531
c	0.000935	2.7	1	2	2	1.45	10	2.4	15000	1324
d	0.000529	0.48	1.5	2	1.83	1.08	11	1.5	14950	1306
e	0.00102	2.7	1	1.85	2	1.33	21	2.4	15120	1279
f	0.002643	2	1	2	1.63	1.13	13	1.5	14895	1221
g	0.00073	0.59	1.5	1	1	1.18	35	1.8	15010	841
h	0.000267	0.5	1	2	1.24	1.31	59	8.1	15000	833
i	0.000233	1.4	1	1	2	1.21	47	7.5	14970	828
l	0.000665	2	1	1.5	1	1.19	38	4.8	15110	691
m	0.000023	0.4	1	1	1	1.23	99	9.9	15040	451
n	0.000537	2	1	1	1.18	1.32	22	3	15060	402
o	0.001665	0	1	1	1	1.18	31	3.6	15105	372
p	0.000466	1	1	1	1	1.34	62	8.4	15000	361
q	0.00005	1	1	1	1	1.1	15	2	14860	300
r	0.00001	1	1	1	1	1.11	16	2.4	14900	269

Furthermore, to establish the danger level for each cluster in terms of accidents, and make it possible to compare them, the hazard level index ( $Id$ , relating to observed accidents) was introduced

$$Id = \frac{(10^8 * N_v * Sev)}{(365 * L * K1 * K2 * ADT)} \quad (10)$$

where:

$N_v$  is the number of vehicles involved in the accident under consideration;

$L$  is the length of the *hazardous zone* (cluster). It was calculated as follows: an influence area (equal to 500m) was assumed for each accident, or accidents, occurring at the same distance.

$K1$  is equal to 0.75 for a dry road surface and 0.25 for a wet road surface (Dell'Acqua et al. 2011b).

$K2$  is equal to 0.67 for daylight and 0.33 for nocturnal light.

$Sev$  is the crash severity. Equal to: 1 no injured; 2 with 1-4 injured; 2.5 with 5-10; 3 with Dead.

$ADT$  is the average daily traffic at each cluster. In particular, the traffic associated with each cluster is the average value for the traffic for each distance where an accident (or accidents) was/were recorded in the "After" period (from 30 September 2002 to 31 August 2003). The traffic for this period was used because the model (11) shown below was used to simulate the accidents for that period (30 September 2002 to 31 August 2003).

This model is only applicable on highways for the values given in the third column of Table 2.

The  $Idc$  was constructed (11) using the data in Table 3, (i.e.  $Id$  as a dependent variable and the variables 1, 2, 3, 4 and 5 are independent variables-Predictors) with a multiple regression equation multivariate analysis (Mauro & Branco, 2013). The ordinary-least-square method was applied to estimate the coefficients of the explanatory variables.

$$Idc = -1415 + 95011CV + 106|LG| + 875EXF + 364LC + 399PV \quad (p2 \ 0.94) \quad (11)$$

The significance of the variables (confidence level < 5%) was examined using the *t-student* test. In Table 4, the last two columns show the values for the *t-student* tests and the significance of the model (11).

Table 4. Model obtained

	Coeff.	Std. Error	<i>t-student</i>	Significance
Constant	-1415.08	185.34	-7.63	0.00
CV	95010.70	42297.01	2.24	0.048
LG	106.63	44.29	2.40	0.037
EXF	875.46	168.90	5.18	0.000
LC	364.42	94.98	3.83	0.003
PV	399.60	108.46	3.68	0.004

#### 4.2 Before-After Analysis

The "Before-After" analysis described in section 2.2 was applied to the first two clusters (cluster "a" and cluster "b") shown in Table 3.

The first cluster (cluster "a"), contains 14 accidents (their average characteristics are shown in the first row of Table 3) in the following kilometric ranges:

4 accidents at km 228.000 ± 0.5 km

5 accidents at km 244.500 ± 0.5 km

5 accidents at km 253.000 ± 0.5 km

The second cluster (cluster "b"), contains 28 accidents (their average characteristics are shown in the second row of Table 3) in the following kilometric ranges:

5 accidents at km 226.000 ± 0.5 km

9 accidents at km 219.500 ± 0.5 km

6 accidents at km  $221.200 \pm 0.5$  km

8 accidents at km  $249.500 \pm 0.5$  km

Bearing in mind the average characteristics of the cluster (i.e., the geometric and environmental situations in which accidents occurred), the following infrastructure projects were planned and implemented.

Cluster “a”  $\rightarrow$  interventions at km  $244.500 \pm 0.5$ :

- 1 Adjustment of acceleration and deceleration lanes at freeway exit.
- 2 Replacing dense asphalt with porous asphalt.

Cluster “b”  $\rightarrow$  interventions at km  $249.000 \pm 0.5$ :

- 1 Increase in curve radius from 390 m to 800 m.
- 2 Replacing dense asphalt with porous asphalt.

Table 5 shows the works proposed to improve road safety in clusters “a” and “b”. The first column shows the area of intervention. The second column shows the *Idc* value for the *after* period but without works (i.e., the “**p**” parameter was established). The third column describes the scheduled works. The fourth column shows the numerical variation of the variables of model (11) after the works in the third column had been planned (Loprencipe & Cantisani, 2013). Finally, the last column shows the *Idc* value for the period after with the work had been carried out (i.e. parameter “**1**” was established).

Table 5. Results obtained by the model (11)

Cluster	Distance of “Hazardous zone”	Hazardous zone index “after” Simulated with model (11) (without work)	Description of the work	Numerical variation of the variable	Hazardous zone index “after” Calculated with model (11) (with work)
A	$244.500 \pm 0.5$	1770	adjustment acceleration and deceleration lanes in Freeway exit	From 1.5 to 1.25	1232
			Changing dense asphalt in Porous asphalt	From 2.0 to 1.20*	
B	$249.000 \pm 0.5$	1374	Curvature: Increase in curve radius from 390 m to 800 m	From 0.00255 to 0.00125	988
			Changing dense asphalt in Porous asphalt	From 2.0 to 1.20*	

\*The variation. as indicated in Table 2. Was should be from 2 to 1. However give that the porous asphalt results in an improvement but not a complete change. From wet t dry conditions. Was associated with a coefficient of 1.20.

Table 6, using the symbols introduced in Section 2.2, shows the results for the estimation of the effectiveness of the modernization works. The first two columns show the values of “**1**” and “**p**” while the fifth and sixth show the values of “ $\theta$ ” and “ $\delta$ ” whose expressions are derived from (6) and (7). The sixth column shows the benefit (as a percentage) of comparison between “**1**” and “**p**”.

As can be seen, the works were effective, resulting  $\theta < 1$ .

Lastly, Table 7 illustrates the “Before-After” comparison at the sites of intervention for the accidents observed. The table shows a comparison (based on the observed data) between the number of accidents, the number of injured, the number of vehicles involved and the traffic before and after the work. The *Observed Hazardous Index (Id)* from Equation (10) is shown. The last two columns show the values of “**1**” and “**p**” calculated from the observed values. As can be seen from a comparison of these two values (“**1**”, “**p**”), in the two situations in Table 6 (for the simulated values) and Table 7 (for the observed values), the estimate carried out using the experimental

procedure are very reliable.

Table 6. Benefits obtained for the accident rate for the simulated “After”, using model (11)

<b>l</b>	<b>p</b>	var( <b>l</b> )	var( <b>p</b> )	<b>Q</b>	<b>d</b>	var( <b>q</b> )	var( <b>d</b> )	Benefits [%]
<b>1232</b>	<b>1770</b>	1232	1770	0,70	<b>538</b>	0.000784	3002	<b>30</b>
<b>988</b>	<b>1374</b>	988	1374	0,72	<b>386</b>	0.001040	2362	<b>28</b>

Table 7. Comparison of the accidents “before and after” observed at the sites of intervention

Distance of Work zone	Number of accidents observed before modernisation works	Number of accidents observed after modernisation works	Number of injured and/or death before modernisation works	Number of injured and/or death after modernisation works	number of damaged vehicles before	number of damaged vehicles After	AADT 1998 - 1999	AADT 2001 - 2002	Before “Hazardous zone index” calculated with Eq. 10	After “Hazardous zone index” calculated with Eq. 10	d (Calculated with observed accidents)	Q (Calculated with observed accidents)	Benefits [%]
244.500 ± 0.5	8	4	4 inj.+ 0 deat.	2 inj.+ 0 deat.	9	8	14670	14920	1505	897	608	0.6	40%
249.500 ± .5	5	2	4 inj.+ 1 deat.	1 inj.+ 0 deat.	7	7	14700	14840	1376	890	486	0.65	35%

## 5. Conclusion

This study has addressed a problem of “Before-After” analysis. Specifically, to carry out this analysis, cluster analysis and multivariate analysis were used. The first technique was used to group incidents characterized by strong similarities and resemblances into the appropriate partitions. The second technique was used to construct a multivariate model (11) to estimate the accident in the “After” configuration under the two conditions “with intervention” and “without intervention”.

The multivariate model (11) was derived from the average characteristics of the accidents contained in each grouping and expressed in terms of a hazard index (10). The two situations (Before and After), simulated by the model, allowed us to apply the “naive before-after” technique. To overcome some of the limitations of this “naïve” technique, and, especially, to make the two situations referred to above comparable, index (10), which makes it possible to relate the traffic rate to the traffic, and the algorithm described in Section 2.1, that allows accidents with similar characteristics to be aggregated in binary groups, were introduced. The results showed the validity of the procedure: through model (11) it was possible to plan and implement the actions listed in Table 6. Moreover, the estimated effectiveness of the “naive Before and After” technique, on the basis of the findings of model (11), was very close to the real effectiveness calculated from the difference between the accidents observed before and after at the modernization worksites (see Tables 6 and 7). Based on the results obtained, a new research plan is being developed where, beyond extending the analysis to an entire road network (consisting of roads of different functional importance) a comparison will be made between all the techniques in the literature for the “Before-After” approach in order to assess the possible benefits of a combination with techniques of the Cluster Analysis type. At the moment, the results obtained from combining cluster analysis with Bayesian analysis seem encouraging. It has been noted in fact that it is very useful to process data using Cluster Analysis since it makes it possible to apply the “Before-After” technique to entities in a meaningful way.

## References

- Dell’Acqua, G. (2012). European speed environment model for highway design-consistency. *Modern Applied Science*, 6(9), 1-10.
- Dell’Acqua, G., De Luca, M., & Lamberti, R. (2011b). Indirect Skid Resistance Measurement for Porous Asphalt Pavement Management. *Transportation Research Record*, 2205, 147-154. <http://dx.doi.org/10.3141/2205-19>
- Dell’Acqua, G., De Luca, M., & Mauro, R. (2011a). Road Safety Knowledge-Based Decision Support System. *Procedia Social and Behavioral Sciences*, 20, 973-983. <http://dx.doi.org/10.1016/j.sbspro.2011.08.106>
- Dell’Acqua, G., De Luca, M., Mauro, R., & Lamberti, R. (2011c). Motorway Speed Management in Southern Italy. *Procedia: Social & Behavioral Sciences*, 20(2011), 49-58.

- De Luca, M., & Dell'Acqua, G. (2012a). Freeway safety management: case studies in Italy. *Transport*, 27(3), 320-326. <http://dx.doi.org/10.3846/16484142.2012.724447>
- De Luca, M., Mauro, R., Lamberti, R., & Dell'Acqua, G. (2012b). Road safety management using Bayesian and cluster analysis. *Procedia: Social & Behavioral Sciences*, 54, 1258-1268.
- De Luca, M., Mauro, R., Russo, F., & Dell'Acqua, G. (2011). Before-After Freeway Accident Analysis using Cluster Algorithms. *Procedia: Social & Behavioral Sciences*, 20(2011), 723-731. <http://dx.doi.org/10.1016/j.sbspro.2011.08.080>
- Depaire, B., Wets, & Vanhoof, K. (2008). Traffic accident segmentation by means of latent class clustering. *Accident Analysis & Prevention*, 40(4), 1257-1266. <http://dx.doi.org/10.1016/j.aap.2008.01.007>
- Ghosh, I. P. T., & Savolainen, T. J. (2012). Gates, Examination of factors affecting freeway incident clearance times: a comparison of the generalized F model and several alternative nested models. *Journal of Advanced Transportation*. <http://dx.doi.org/10.1002/atr.1189>
- Hauer, E. (1997). Observational Before-After Studies in Road Safety. Pergamon Publication. Oxford, UK: Elsevier Science.
- Kwok-suen, N., Wing-tat, H., & Wing-gun, W. (2002). An algorithm for assessing the risk of traffic accident. *Journal of Safety Research*, 33(3), 387-410. [http://dx.doi.org/10.1016/S0022-4375\(02\)00033-6](http://dx.doi.org/10.1016/S0022-4375(02)00033-6)
- Loprencipe, G., & Cantisani, G. (2013). Unified Analysis of Road Pavement Profiles for Evaluation of Surface Characteristics. *Modern Applied Science*, 7(8), 1-14. <http://dx.doi.org/10.5539/mas.v7n8p1>
- Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, 44(5), 291-305. <http://dx.doi.org/10.1016/j.tra.2010.02.001>
- Mauro, R., & Branco, F. (2012). Two Vehicular Headways Time Dichotomic Models. *Modern Applied Science*, 6(12), 1-12. <http://dx.doi.org/10.1016/j.tra.2010.02.001>
- Tryon, R. C. (1939). *Cluster analysis*. New York: McGraw-Hill.

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).