

# ASPECTELE REGIONALE ALE CULTURII CUCUTENI FAZA A DIN PERSPECTIVĂ STATISTICĂ. TEORIE ȘI STUDIU DE CAZ ASUPRA VARIAȚIEI FORMEI PAHARELOR DIN CINCI AȘEZĂRI

DE  
GEORGE BODI\*

**Abstract:**

*Our study investigates whether, during the phase A of the Cucuteni culture, a connection exists between observed regional cultural preferences for pottery decoration and pottery shapes. We do so through the statistical investigation of the variation in the shape of beakers, expressed as the ration between maximum diameter and height, from five diagnostic settlements. The use of one way ANOVA allows us to ascertain that there is significant variation between settlements. A post hoc Scheffe test, comparing the pairwise variability of the variation, shows that the specific decoration characteristics used to define each cultural group are also accompanied by preferences in specific beaker shapes. We conclude that, although our data set forces caution on the generalisation of our results, shape variation may be used as an added argument in the identification and characterisation of regional groups.*

**Keywords:** *Cucuteni A cultural groups, beakers, shape variation, one way ANOVA, post hoc Scheffe test.*

## INTRODUCERE

Existența unor aspecte culturale regionale în cadrul fazei Cucuteni A a fost avansată pentru prima dată de Vladimir Dumitrescu în 1974, definindu-se la acea dată două tradiții diferite de decorare a ceramicii, pornind de la două dintre cele mai bine cunoscute situri ale perioadei: Hăbășești și Drăgușeni<sup>1</sup>. În cadrul aceleiași discuții sunt observate și caracteristicile deosebite ale materialelor provenite din așezările de la Fedeleșeni și Târpești, fără însă ca acestea să fie considerate ca reprezentative pentru fenomene regionale clar definibile. Timp de aproape trei decenii, asupra acestei probleme se revine rar și fără o re-aducere la zi a conceptelor propuse din perspectiva descoperirilor ulterioare<sup>2</sup>. Începutul anilor 2000 schimbă această situație, aducând nu numai monografia dedicată aspectului cultural Drăgușeni<sup>3</sup> dar și o re-evaluare a caracteristicilor regionale identificate de Vladimir Dumitrescu. Pentru faza A, Dumitru Boghian reconfirmă existența individualității decorului ceramic din așezările de la Hăbășești și Drăgușeni, și delimitează, în mod clar, noi entități regionale precum aspectul Horodnica-Trușești-Cuconești Vechi, aspectul cu benzi înguste (în care este inclusă și așezarea de la Fedeleșeni) și aspectul subcarpatic, cuprinzând situri precum cele de la Poduri, Frumușica, Izvoare sau Ghelăiești<sup>4</sup>.

În acest punct, trebuie să observăm că, deși studierea calitativă a caracteristicilor decorului ceramic joacă un rol major în ordonarea crono-spațială a descoperirilor provenind din spațiul cucutenian, o mare parte din rezultatele cercetărilor arheologice prezintă o importantă componentă cantitativă, și totuși, utilizarea

---

\* Institutul de Arheologie, Academia Română-Filiala Iași; georgebodi@gmail.com.

<sup>1</sup> DUMITRESCU 1974.

<sup>2</sup> MARINESCU-BÎLCU 1991; PETRESCU-DÎMBOVIȚA 1996; MANTU 1998: 71-76.

<sup>3</sup> SOROCHIN 2002.

<sup>4</sup> BOGHIAN 2001.

analizelor statistice a seturilor de date obținute cu scopul revelării unor structuri discrete constituie o abordare mult prea rar întâlnită. În cazul cercetării culturii Cucuteni, analizele construite pe informații cantitative și care reușesc să depășească prezentarea unor simple comparații numerice sau procentuale sunt foarte rare, fiind vizată în special analiza ceramicii<sup>5</sup> și, mai nou, a utilajului litic<sup>6</sup>.

Pornind de la cele două observații din rândurile de mai sus, scopul articolului nostru este unul dublu. Într-o primă instanță vom încerca să observăm dacă preferințele regionale pentru un anumit tip de decorare a ceramicii specifice fazei Cucuteni A se reflectă și în variații ale formei. În al doilea rând, vom utiliza această ocazie pentru prezentarea unui flux complet de analiză a unui set de date cu ajutorul unor tehnici statistice univariate de bază.

## CONCEPTE ȘI METODELE ALE CERCETĂRII

Pentru transpunerea variabilității formelor ceramice într-un set de date convenabil pentru analiză statistică cu ajutorul unor tehnici univariate, am să ales să utilizăm exprimarea caracteristicilor formei ca raport dintre diametrul maxim și înălțime, această măsură fiind utilizată cu succes în construirea de comparații între forme de vase provenite din seturi etnografice cu forme de vase din colecții arheologice, în scopul determinării funcționalității celor din urmă<sup>7</sup>. Pentru analiza noastră am ales un set de așezări reprezentative, care au beneficiat de publicarea rezultatelor cercetării în formă monografică, cu potențial de a ne furniza seturi de date de dimensiuni suficiente pentru o analiză statistică. Am utilizat astfel monografiile de la Hăbășești<sup>8</sup>, Trușești<sup>9</sup>, Târpești<sup>10</sup>, și Drăgușeni<sup>11</sup> pentru ilustrarea aspectelor culturale cu același nume, precum și monografiile de la Frumușica<sup>12</sup> pentru aspectul subcarpatic și Preutești<sup>13</sup>, ca posibilă confirmare a datelor provenite din așezarea de la Hăbășești. Forma aleasă pentru studiu este cea a paharelor, deoarece ne-a permis culegerea unui set de 232 de observații, un număr suficient pentru scopul nostru.

*Tipuri de variabile.* În tabelul 1, care prezintă șase cazuri extrase aleatoriu, se poate observa că în setul nostru de date avem două coloane, pe care le vom denumi în continuare variabile, deoarece reprezintă caracteristici individuale ale paharelor măsurate. Prima variabilă conține coduri pentru identificarea așezării de proveniență a fiecărui pahar măsurat (ex. Tr indică așezarea de la Trușești). Pentru că există un număr limitat de valori pe care această variabilă poate să le ia (în cazul nostru doar șase), această variabilă este denumită categorică. Pentru că valorile din această variabilă nu depind de nici o alta, aceasta este în același timp și o *variabilă independentă*. Cea de a doua variabilă conține valorile raportului dintre diametrul maxim și înălțimea paharelor analizate. Pentru că este exprimată numeric și poate lua un număr infinit de valori, această variabilă este denumită discretă. Pentru că valorile acestei variabile depind de coloana care conține numele așezărilor, aceasta este de asemenea și o *variabilă dependentă*<sup>14</sup>.

---

<sup>5</sup> LAZAROVICI, LAZAROVICI 2012: 209-289; MAXIM-KALMAR, TARCEA 1999.

<sup>6</sup> VORNICU 2017.

<sup>7</sup> HENRICKSON, MCDONALD 1983.

<sup>8</sup> DUMITRESCU *et al.* 1954.

<sup>9</sup> PETRESCU-DÎMBOVIȚA *et al.* 1999.

<sup>10</sup> MARINESCU-BÎLCU 1981.

<sup>11</sup> MARINESCU-BÎLCU, BOLOMEY 2000.

<sup>12</sup> MATASĂ 1946.

<sup>13</sup> URSULESCU, IGNĂTESCU 2003.

<sup>14</sup> HEUMANN, SCHOMAKER, SHALABH 2016: 6.

**Tabel 1.** Extras din setul de date.

Așezare	Raport
Tr	1.10
Ha	1.15
Ha	1.00
Dr	1.22
Tr	1.14
Fr	1.01

*Analiza exploratorie* constituie primul pas în crearea unei imagini sintetice asupra setului de date analizat. În această etapă vom urmări trei seturi de informații distincte: măsurile tendinței centrale, ale variației și ale distribuției.

*Măsurile tendinței centrale* sunt constituite de valorile care indică centrul setului de date și sunt de mai multe tipuri. În cazul studiului nostru, de interes sunt media și mediana. Media (Med) reprezintă media aritmetică a cărei valoare este obținută după formula Suma/Nr. Mediana (Mdn) este valoarea care împarte setul de date în două jumătăți egale<sup>15</sup>.

*Măsuri ale variației.* O primă măsură a variației este constituită de abaterea standard (as). Aceasta se calculează prin extragerea rădăcinii pătrate din suma diferențelor individuale față de medie, ridicate la pătrat. Această valoare ne indică măsura în care valorile individuale ale unui variabile variază față de medie.<sup>16</sup> Pornind de la abaterea standard, un al doilea concept la care ne vom raporta este cel al erorii standard. Aceasta se calculează împărțind abaterea standard la rădăcina pătrată a numărului de observații. Valoarea astfel obținută constituie un indicator al gradului de încredere în media variabilei analizate: cu cât eroarea standard este mai mică, cu atât este mai puțin probabil ca media să se modifice prin introducerea de noi observații<sup>17</sup>. Amplitudinea unei variabile se calculează ca diferența dintre valoarea maximă și valoarea minimă și oferă o imagine sintetică a întinderii acesteia<sup>18</sup>. Quantilele sunt puncte ce separă setul de valori al unei variabile în părți egale. Cele mai utilizate, și pe care le vom folosi și noi, sunt denumite și quartile. Primul quartil separă cele mai joase 25% dintre valori de restul de 75%. Al doilea quartil este constituit de mediană și împarte setul de date în două părți egale. Cel de al treilea quartil separă cele mai înalte 25% dintre valori de restul de 75%. Amplitudinea interquartilă, calculată ca diferența dintre al treilea și primul quartil, constituie o măsură a variabilității mai puțin sensibilă la prezența valorilor extreme decât devierea standard<sup>19</sup>. Valorile extreme sunt constituite de puncte specifice cu valori care nu se potrivesc cu restul datelor. În cazul analizelor univariate, ca regulă generală, o valoare este considerată ca fiind extremă dacă se află față de primul sau al treilea quartil la o distanță de 1.5 ori mai mare decât distanța interquartilă<sup>20</sup>. Identificarea valorilor extreme este importantă deoarece pot influența validitatea rezultatelor analizei datelor. Pentru acest studiu de caz, am considerat potrivit să calculăm și intervalul de încredere de 95% pentru valorile mediei fiecărui grup, care ne va permite să testăm dacă există eventuale diferențe observabile în estimarea acestora<sup>21</sup>.

Cunoașterea *măsurilor distribuției* este importantă în special pentru determinarea tipului de analiză statistică ce urmează a fi utilizat. În cazul unei distribuții normale, care din punct de vedere grafic poate fi reprezentată ca o curbă Gauss, valoarea indicelui de asimetrie este 0, iar a indicelui de aplatizare este 3, iar seturile de date pot fi analizate utilizând tehnici statistice parametrice. Totuși, valorile de 0 și respectiv 3 sunt

<sup>15</sup> MACFARLAND 2014: 7.

<sup>16</sup> DENIS 2016: 109.

<sup>17</sup> DUNN 2001: 328.

<sup>18</sup> HAN, KAMBER, PEI 2011: 48.

<sup>19</sup> PECK, OLSEN, DEVORE 2016: 167.

<sup>20</sup> KERNS 2018: 44.

<sup>21</sup> HECTOR 2015: 73; CAMPBELL, SWINSCOW 2009: 50-52.

rare în cazul datelor provenite din lumea reală. Astfel, în cazul indicelui de asimetrie, valori cuprinse între 0.5 și -0.5 sunt considerate ca indicând o asimetrie ușoară, cele între 0.5 și 1 și -0.5 și -1 o asimetrie medie, iar valorile mai mari decât 1 sau -1 indică asimetrii puternice. Valori ale indicelui de aplatizare mai mari decât 3 caracterizează o distribuție leptokurtică, cu valori concentrate în jurul mediei și cu o probabilitate mai mare de apariție a valorilor extreme. Valori mai mici decât 3 sunt specifice distribuțiilor platykurtice, cu distribuție mai largă în jurul mediei, și cu o probabilitate mai mică de apariție a valorilor extreme<sup>22</sup>. Pentru a evita, pe cât posibil, erorile de judecată în stabilirea tipului de distribuție al unui set de valori, pentru un număr de observații mai mic sau apropiat de 30 se recomandă utilizarea unor teste statistice de verificare a normalității distribuției, în timp ce pentru seturi cu observații care depășesc cu mult 30 de cazuri, putem presupune normalitatea distribuției în virtutea examinării histogramei de distribuție a valorilor.

În cazul nostru, pentru confirmarea normalității distribuției fiecărui set de valori provenit din așezările analizate, vom utiliza testul Shapiro-Wilk. Acesta compară valorile setului analizat cu valorile specifice unui set cu distribuție normală care prezintă aceeași medie și deviere standard<sup>23</sup> pornind de la ipoteza că mulțimea analizată este caracterizată de o distribuție normală.

Pentru că în studiul nostru ne dorim compararea unor grupuri de valori, pentru a asigura validitatea rezultatelor, o altă condiție preliminară care trebuie îndeplinită este existența unei variabilități omogene față de medie între grupuri. Pentru stabilirea acesteia vom utiliza testul Levene, care compară variația dintre diferența valorii măsurate a fiecărui caz a grupului față de media acestuia, pornind de la ipoteza că variația între grupuri este omogenă<sup>24</sup>.

În cazul ambelor teste, rezultatul este exprimat ca valoare a unei probabilități care permite acceptarea sau respingerea ipotezei testate. În cazul nostru, valoarea probabilității (cunoscută și ca valoarea  $p$ ) cu care vom lucra este de 0,05, care înseamnă că există o probabilitate de 95% ca ipoteza testată să fie relevantă și, drept urmare, poate fi acceptată ca validă<sup>25</sup>.

Diferențele dintre grupuri vor fi testate din punctul de vedere al relevanței statistice cu ajutorul analizei variației unidirecționale (*one way ANOVA*) deoarece setul nostru de date conține o variabilă categorică independentă, care definește șase grupuri și față de care analizăm o variabilă discretă dependentă. ANOVA este o procedură statistică inferențială utilizată pentru identificarea diferențelor mediilor între sau în interiorul unor grupuri de date pornind de la împărțirea variației calculate în interiorul fiecărui grup la variația calculată între grupuri. Rezultatul obținut este denumit valoare  $F$ , pentru care se calculează apoi valoarea  $p$  testându-se ipoteza conform căreia grupurile nu prezintă diferențe semnificative între medii<sup>26</sup>.

În cazul în care analiza variației identifică existența unor diferențe semnificative statistic între grupuri, analiza trebuie continuată cu un test *post-hoc*, capabil să identifice perechile de grupuri specifice care prezintă diferențe semnificative ale mediilor. În cazul nostru vom folosi metoda Scheffe, deoarece este adaptată și comparațiilor între grupuri cu mulțimi de observații inegale, așa cum este și cazul nostru<sup>27</sup>.

Analiza statistică a setului nostru de date o vom desfășura cu ajutorul soluției *open-source* oferite de mediul și limbajul R<sup>28</sup>, în forma oferită de mediul de dezvoltare integrat al RStudio<sup>29</sup>. Pentru efectuarea analizelor statistice descrise în rândurile de mai sus, în afara distribuției de bază, am adăugat și utilizat următoarele pachete:

`readxl`<sup>30</sup> – pentru importarea datelor salvate în format excel;

---

<sup>22</sup> CANNING 2013: 48-49

<sup>23</sup> FIELD 2009: 144.

<sup>24</sup> FIELD 2009: 150.

<sup>25</sup> DUNN 2001: 306.

<sup>26</sup> CANNING 2013: 113-117.

<sup>27</sup> MILLER 1986: 74; JUDD, MCCLELLAND, RYAN 2017: 200.

<sup>28</sup> R CORE TEAM 2019.

<sup>29</sup> RSTUDIO TEAM 2016.

<sup>30</sup> WICKHAM, BRYAN 2019.

psych<sup>31</sup> și Rmisc<sup>32</sup> – pentru obținerea descrierilor statistice sintetice;  
 ggplot2<sup>33</sup> și lattice<sup>34</sup> – pentru vizualizarea unora dintre datele și rezultatele analizelor;  
 MVN<sup>35</sup> – pentru efectuarea analizei normalității distribuției pe grupuri;  
 car<sup>36</sup> – pentru analiza omogenității variației mediei între grupuri;  
 agricolae<sup>37</sup> – pentru efectuarea comparației diferențelor între medii prin metoda Scheffe.

Rezultatele textuale și tabelare prezentate în cadrul acestui articol au fost sintetizate și adaptate pentru forma publicabilă a materialului din rezultatele extinse oferite de soluția noastră software. Analiza noastră poate fi replicată și/sau verificată accesând anexele care însoțesc forma electronică a articolului pe pagina web a periodicului. Anexa I conține codul R explicat pentru replicarea analizelor, precum și rezultatele originale extinse, în format .Rmd. Anexa II conține tabelul excel cu datele utilizate în analiza noastră.

## PREZENTAREA ȘI ANALIZA DATELOR

Tabelul 2 prezintă o serie de valori care caracterizează în mod sintetic setul de date analizat, rezultatele fiind prezentate grupat, în funcție de așezările de proveniență. Un prim aspect care se observă constă în numărul mic de cazuri (doar patru), care provin din așezarea de la Preutești și care sunt caracterizate prin cele mai mari valori ale erorii standard și intervalului de încredere, indicând o variabilitate mare a raportului dintre diametrul maxim și înălțime, în ciuda numărului mic de cazuri. În rest, în cazul nostru, examinarea tabelului se dovedește a fi prea puțin informativă astfel încât, pentru crearea unei impresii generale asupra caracteristicilor formelor paharelor avem nevoie de sprijin grafic.

**Tabel 2.** Valori ale statisticilor descriptive ale setului de date analizat.

Așezare	Nr. Obs.	Med	as	Mdn	Min.	Max.	Ampl.	1 Qu.	3 Qu.	Ind. Asim.	Ind. Aplat.	Er.St.	CI 95%
Drăgușeni	32	1	0.12	1	1	1.29	0.61	0.9350	1.0525	-0.03	0.51	0.02	±0.044
Frumușica	37	1.08	0.08	1.07	1.07	1.29	0.35	1.040	1.110	0.68	0.27	0.01	±0.025
Hăbășești	42	1.15	0.09	1.15	1.15	1.33	0.42	1.090	1.210	-0.27	-0.45	0.01	±0.030
Preutești	4	1.07	0.08	1.11	1.11	1.13	0.18	1.062	1.123	-0.7	-1.72	0.04	±0.134
Târpești	22	1.1	0.12	1.08	1.08	1.32	0.52	1.042	1.160	-0.22	-0.05	0.03	±0.055
Trușești	95	1.08	0.1	1.09	1.09	1.43	0.67	1.030	1.140	-0.25	1.44	0.01	±0.020

Figura 1 este o diagramă boxplot care ne oferă o primă vizualizare a modului de dispersie a datelor în funcție de mediană. Putem astfel să observăm că există o diferență clară între forma paharelor provenite din așezările de la Drăgușeni și Hăbășești, pentru care nu există suprapuneri ale valorilor situate în intervalul interquartil. Așezările de la Preutești, Frumușica, Trușești și Târpești se situează între cele două cazuri extreme, fără a prezenta diferențe notabile una față de cealaltă.

<sup>31</sup> REVELLE 2018.

<sup>32</sup> HOPE 2013.

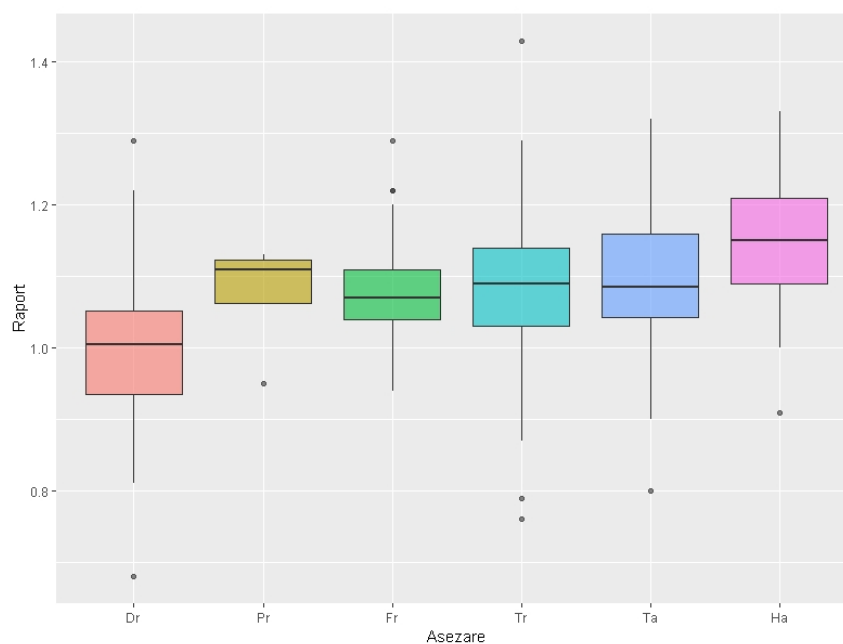
<sup>33</sup> WICKHAM 2016.

<sup>34</sup> SARKAR 2008.

<sup>35</sup> KORKMAZ, GOKSULUK, ZARARSIZ 2014.

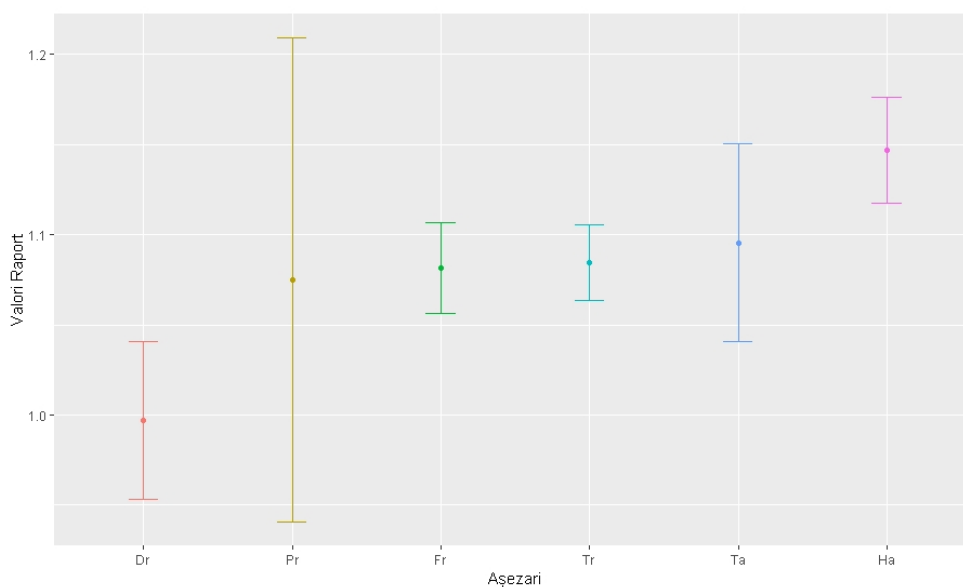
<sup>36</sup> FOX, WEISBERG 2011.

<sup>37</sup> DE MENDIBURU 2019.



**Fig. 1.** Boxplot cu distribuția valorilor grupurilor analizate în funcție de mediană și intervalul interquartil (Dr=Drăgușeni; Pr=Preutești; Fr=Frumușica; Tr=Trușești; Ta=Târpești; Ha=Hăbășești).

Figura 2 reprezintă grafic intervalele de încredere de 95%, care ne ajută să vizualizăm probabilitățile de situație pentru media valorilor specifice fiecărei așezări. În acest caz, observăm că se conturează trei grupări distincte alcătuite din așezarea de la Drăgușeni, un grup, Hăbășești, al doilea grup, și Frumușica și Trușești, al treilea grup, care prezintă o probabilitate mai mică de 5% ca valoarea mediei să se situeze în interiorul unui interval comun. În cazul așezării de la Preutești, dispersia mare a intervalului este un rezultat direct al numărului mic de observații, neexistând suficiente date pentru o aproximare mai precisă. În cazul așezării de la Târpești observăm că aceasta are puncte comune atât cu Frumușica și Trușești cât și cu Hăbășești.

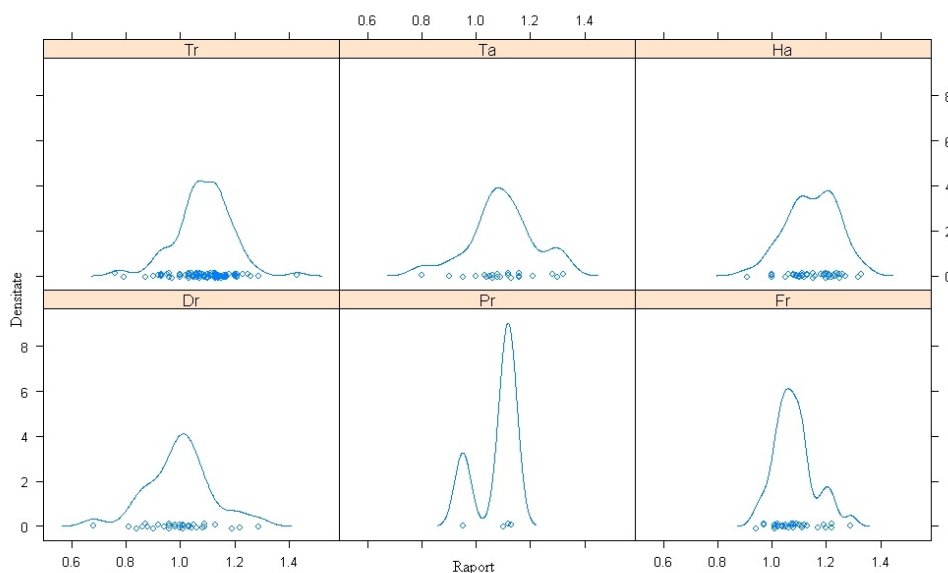


**Fig. 2.** Distribuția probabilităților situării mediilor grupurilor analizate cu un interval de încredere de 95%.

Observațiile de până acum ne permit să concluzionăm că există o probabilitate de cel puțin de 95% ca formele paharelor provenite din așezările studiate să fie caracterizate de forme diferite exprimate de raportul dintre diametrul maxim și înălțime. În continuare, cu ajutorul ANOVA, dorim să verificăm dacă aceste diferențe, așa cum sunt acestea vizibile în intervalele de probabilitate ale mediei, sunt confirmate și de analiza diferenței variabilității existente în cadrul măsurătorilor specifice fiecărei așezări față de variabilitatea existentă între așezări. Pentru utilizarea acestui test, trebuie însă să ne asigurăm că sunt îndeplinite condițiile de normalitate și de variație omogenă a distribuției datelor. Pentru primul caz, valorile mici ale indicelui de asimetrie din tabelul 2 ne indică o distribuție cu o asimetrie medie, dar valorile indicelui de applatizare ne indică prezența unor distribuții platykurtice. Verificarea normalității distribuției datelor cu ajutorul testului Shapiro-Wilks (Tabel 3) și a inspecției graficelor de distribuție a densității (Fig. 3) ne arată că putem considera seturile de date provenite din așezările de la Drăgușeni, Frumușica, Hăbășești și Târpești ca prezentând distribuții normale. Deși testul Shapiro-Wilks respinge ipoteza unei distribuții normale în cazul așezării de la Trușești, noi o vom accepta în baza inspecției vizuale a histogramei. În ceea ce privește așezarea de la Preutești, acceptăm respingerea ipotezei de normalitate în baza testului și a distribuției bimodale prezente în grafic. Dat fiind și numărul mic de observații din acest sit, considerăm că aceste date nu ar putea contribui în mod semnificativ la îmbunătățirea puterii de analiză, fiind posibil chiar să adauge zgomot de fond și să distorsioneze rezultatele următoare, astfel încât, în continuare, vom renunța la utilizarea celor patru măsurători.

**Tabel 3.** Rezultatele testului Shapiro-Wilks pentru normalitatea distribuției.

Așezare	Nr. Obs.	Valoare F	Valoare p	Normalitate
Drăgușeni	32	0.9786801	0.7603894	DA
Frumușica	37	0.954128	0.1310867	DA
Hăbășești	42	0.9794636	0.6408369	DA
Preutești	4	0.7576391	0.04541828	NU
Târpești	22	0.968886	0.6852827	DA
Trușești	95	0.9723423	0.04162913	NU



**Fig. 3.** Distribuția densității valorilor grupurilor analizate.

Rezultatul testului Levene (Tabel 4) cu valoarea  $p > 0.05$  ne permite să acceptăm ipoteza conform căreia variația între cele cinci așezări rămase în analiză este omogenă. Condițiile de bază pentru ANOVA fiind

îndeplinite, putem efectua analiza variației, rezultatele acesteia fiind prezentate în tabelul 5. Valoarea p, puternic negativă, ne indică faptul că există diferențe semnificative statistic între dimensiunile paharelor specifice fiecărei așezări. Această constatare este confirmată și de valoarea F, care este obținută prin împărțirea variației dintre grupuri la cea specifică fiecărui grup<sup>38</sup>, și a cărei valoare, în cazul nostru, indică faptul că variația dintre grupuri este de aproape 10 ori mai mare decât cea din interiorul grupurilor.

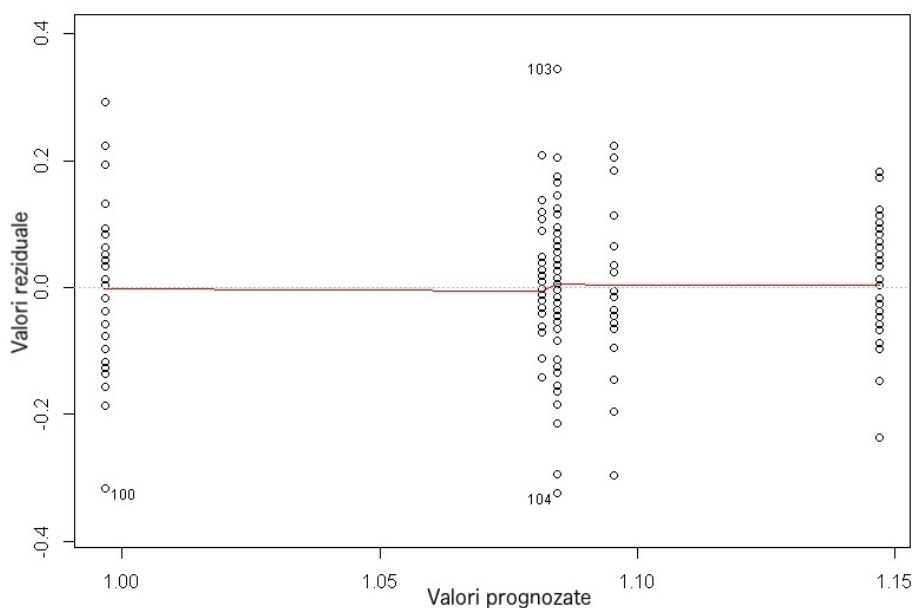
**Tabel 4.** Rezultatele testului Levene pentru omogenitatea variației.

Test Levene pentru omogenitatea variației (centru = mediana)			
	Grade de libertate	Valoare F	Valoare p
grupuri	4	1.3581	0.2495
cazuri	223		

**Tabel 5.** Rezultatele analizei variației.

	Grade de libertate	Suma pătratelor	Media pătratelor	Valoare F	Valoare p
Asezare	4	0.4123	0.10306	9.844	2.37x10 <sup>-7</sup>
Reziduale	223				

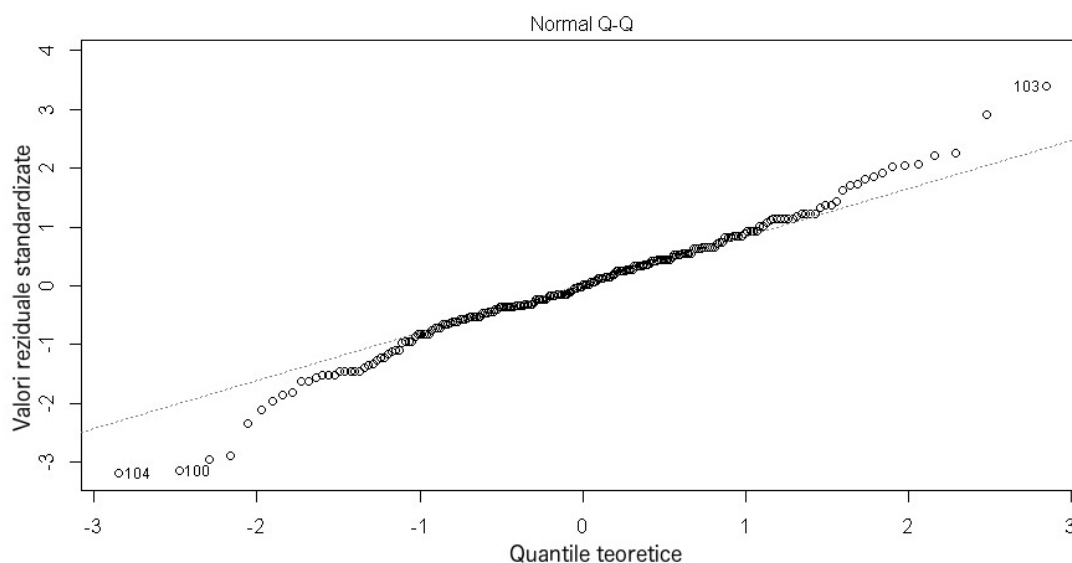
Validitatea analizei ne este confirmată și de diagrama de distribuție a rezidualelor față de linia valorilor medii, unde nu se observă o relație evidentă între acestea (Fig. 4), precum și de acordul observabil în distribuția rezidualelor față de linia ipotetică a unei distribuții normale (Fig. 5).



**Fig. 4.** Diagrama de distribuție a rezidualelor față de valorile prognozate în urma ANOVA.

<sup>38</sup> DUNN 2001: 414-418.





**Fig. 5.** Grafic de distribuție a rezidualelor față de linia ipotetică a unei distribuții normale în urma ANOVA.

Cele două grafice ale distribuției rezidualelor ne semnalează, de asemenea, prezența a trei valori extreme situate în setul nostru de date la rândurile 100, 103 și 104. Deoarece nu putem să le atribuim unor greșeli de colectare a datelor, acceptăm faptul că aceste trei cazuri nu pot fi eliminate din analiză. Acțiuni standard recomandate pentru a încerca să le micșorăm impactul asupra analizei noastre constau în transformarea logaritmică (Tabel 6) sau în aducerea acestora la scară în urma scăderii valorii mediei din fiecare măsurătoare și împărțirea rezultatului la valoarea deviației standard (Tabel 7)<sup>39</sup>.

**Tabel 6.** Rezultatele analizei variației cu valorile setului de date transformate logaritmic în bază naturală.

	Grade de libertate	Suma pătratelor	Media pătratelor	Valoare F	Valoare p
Asezare	4	0.3815	0.09538	10.05	1. x10 <sup>-7</sup>
Reziduale	223	2.1173	0.00949		

**Tabel 7.** Rezultatele analizei variației cu valorile setului de date aduse la scară.

	Grade de libertate	Suma pătratelor	Media pătratelor	Valoare F	Valoare p
Asezare	4	34.07	8.517	9.844	2.37x10 <sup>-7</sup>
Reziduale	223	192.93	0.865		

Repetarea analizei cu datele transformate prin cele două metode ne permite să observăm că rezultatele nu sunt cu mult îmbunătățite, de unde deducem că impactul acestor valori nu este semnificativ, astfel încât ne putem continua analiza cu setul de date netransformate. Graficele de verificare a rezidualelor pentru analiza variației datelor transformate pot fi reproduse prin informația conținută în format electronic de Anexele 1 și 2.

Testul *post hoc* pentru analiza variației pe perechi de grupuri, prezentat în tabelul 8, ne confirmă imaginea construită de vizualizarea intervalelor de încredere. În acest moment putem afirma faptul că forma paharelor din așezările analizate cunoaște diferențe statistice puternic semnificative. Astfel, așezarea de la Hăbășești este caracterizată de o medie a raportului de 1.14 cu o abatere standard de 0.09, exprimând o preferință pentru pahare mai largi și mai puțin înalte. Pentru așezările de la Trușești (media = 1.08, abatere

<sup>39</sup> TABACHNICK, FIDELL 2013: 77.

standard = 0.1) și Frumușica (media = 1.08, abatere standard = 0.07) sunt specifice, de asemenea, paharele cu diametrul maxim mai mare decât înălțimea, dar raportul preferat dintre aceste două dimensiuni este mai mic decât la Hăbășești, media acestuia fiind de 1.08 în ambele cazuri. Așezarea de la Drăgușeni se distinge printr-o formă a paharelor în care diametrul maxim este aproximativ egal cu înălțimea, media raportului dintre aceste două dimensiuni fiind de 0.99, iar abaterea standard de 0.1. Forma paharelor provenite de la Târpești prezintă caracteristici comune celor două grupuri definite de locuirile cucuteniene de fază A de la Hăbășești, dar și Trușești și Frumușica, cu o medie a raportului dintre diametrul maxim și înălțime de 1.09 și abaterea standard de 0.1.

**Tabel 8.** Rezultatele testului Scheffe.

Eroarea minimă la pătrat: 0.01046958					
Medii pe așezări					
	Raport	Abatere standard	Nr. Obs	Minim	Maxim
Drăgușeni	0.996875	0.12169336	32	0.68	1.29
Frumușica	1.081351	0.07528325	37	0.94	1.29
Hăbășești	1.146905	0.09369093	42	0.91	1.33
Târpești	1.095455	0.12393267	22	0.80	1.32
Trușești	1.084526	0.10258124	95	0.76	1.43
Alfa: 0.05		Eroare grade de libertate: 223			
Valoare F critică: 2.412129					
Grupuri potrivit probabilității diferențelor mediilor și nivelului alfa (0.05)					
Mediile cu aceeași literă nu sunt diferite semnificativ					
	Raport	Grup			
Hăbășești	1.146905	<b>a</b>			
Târpești	1.095455	<b>ab</b>			
Trușești	1.084526	<b>b</b>			
Frumușica	1.081351	<b>b</b>			
Drăgușeni	0.996875	<b>c</b>			

## CONCLUZII

Pentru a rezuma, observăm cum analiza statistică exploratorie a setului de date analizat ne-a indicat prezența a patru grupuri distincte în preferințele de modelare a paharelor dintr-o serie de așezări cucuteniene de fază A, în funcție de raportul dintre diametrul maxim și înălțime. Aceste observații au fost confirmate prin analize statistice inferențiale, care au testat ipoteze referitoare la variabilitatea mediilor măsurătorilor efectuate și care au stabilit faptul că diferențele observabile în variația formei paharelor între așezările studiate prezintă o probabilitate mai mică de 5% de a constitui rezultatul hazardului. Astfel, studiul nostru reușește să completeze și să consolideze considerațiile privitoare la existența unor aspecte culturale distincte în cadrul fazei A a culturii Cucuteni. Diferențele de decorare a vaselor observate între așezările de la Hăbășești, Trușești, Frumușica și Drăgușeni sunt acum confirmate statistic și de analiza formei paharelor. Aparenta identitate a formelor paharelor din așezările de la Trușești și Frumușica este, din punctul nostru de vedere, rezultatul transmiterii unor tradiții culturale observabile și în cadrul modului de decorare a ceramicii<sup>40</sup>. În baza aceluiași criteriu se confirmă, de asemenea, și caracterul mixt al tradițiilor observate în cazul așezării de la Târpești. În același timp, trebuie să atragem atenția că natura procedurilor statistice utilizate și a datelor folosite ne îndeamnă la precauție în ceea ce privește puterea generalizatoare a concluziilor obținute. Astfel, pentru noi, rezultatele obținute sunt valide doar pentru așezările și forma ceramică analizată, dar, pentru cercetările viitoare, remarcăm faptul că acest tip de demers poate constitui un instrument de lucru ce deschide noi posibilități de analiză a variațiilor culturale

<sup>40</sup> BOGHIAN 2001: 95.

observabile prin studierea decorului ceramic, susținându-le printr-un set de argumente independente provenit din investigarea preferințelor în modelarea formelor vaselor.

## BIBLIOGRAFIE

- BOGHIAN 2001 Boghian, D., *Quelques considerations sur les aspects regionaux du complexe Cucuteni-Tripolye*, in: *Memoria Antiquitatis*, XXII, 71-114.
- CAMPBELL, SWINSCOW 2009 Campbell, M. J., Swinscow, T. D. V., *Statistics at square one*, Wiley-Blackwell, Oxford.
- CANNING 2013 Canning, J., *Statistics for the Humanities*, British Academy, London.
- DE MENDIBURU 2019 de Mendiburu, F., *Agricolae: Statistical Procedures for Agricultural Research*, in: The R Project for Statistical Computing <https://CRAN.R-project.org/package=agricolae> (14.03.2019).
- DENIS 2016 Denis, D. J., 2016. *Applied univariate, bivariate, and multivariate statistics*, John Wiley & Sons, Inc, Hoboken.
- DUMITRESCU 1974 Dumitrescu, Vl., *Aspecte regionale în aria de răspândire a culturii Cucuteni, în cursul primei sale faze de dezvoltare*, in: *Studii și Cercetări de Istorie Veche și Arheologie*, 25, 4, 545-554.
- DUMITRESCU et al. 1954 Dumitrescu, Vl., Dumitrescu, H., Petrescu-Dîmbovița, M., Gostar, N., *Hăbășești: monografie arheologică*, Ed. Academiei R.P.R., București.
- DUNN 2001 Dunn, D. S., *Statistics and data analysis for the behavioral sciences*, McGraw-Hill, New York.
- FIELD 2009 Field, A., *Discovering Statistics Using SPSS*, Sage Publications Ltd., London.
- FOX, WEISBERG 2011 Fox, J., Weisberg S., *An {R} Companion to Applied Regression*, Sage, Thousand Oaks.
- HAN, KAMBER, PEI 2011 Han, J., Kamber, M., Pei, J., *Data Mining. Concepts and Techniques*, Elsevier, Waltham.
- HECTOR 2015 Hector, A., *The New Statistics with R. An Introduction for Biologists*, Oxford University Press, Oxford.
- HENRICKSON, MCDONALD 1983 Henrickson, E. F., McDonald, M. M. A., *Ceramic Form and Function: An Ethnographic Search and an Archeological Application*, in: *American Anthropologist*, New Series, 85, 3, 630-643.
- HEUMANN, SCHOMAKER, SHALABH 2016 Heumann, C., Schomaker, M., Shalabh, *Introduction to Statistics and Data Analysis. With Exercises, Solutions and Applications in R*, Springer, Cham.
- HOPE 2013 Hope, R. M., *Rmisc: Ryan Miscellaneous*, in: The R Project for Statistical Computing, <https://CRAN.R-project.org/package=Rmisc> (14.03.2019).
- JUDD, MCCLELLAND, RYAN 2017 Judd, C. M., McClelland, G. H., Ryan, C. S., *Data analysis: a model comparison approach to regression, ANOVA, and beyond*, Routledge, New York.
- KERNS 2018 Kerns, J. G., *Introduction to Probability and Statistics Using R*, in: The Comprehensive R Archive Network, <https://cran.r-project.org/web/packages/IPSUR/vignettes/IPSUR.pdf> (14.03.2019).
- KORKMAZ, GOKSULUK, ZARARSIZ 2014 Korkmaz, S., Goksuluk, D., Zararsiz, G., *MVN: An R Package for Assessing Multivariate Normality*, in: *The R Journal*, 6, 2, 151-162.
- LAZAROVICI, LAZAROVICI 2012 Lazarovici, C.-M., Lazarovici, Gh., *Ruginoasa – Dealul Drăghici: monografie arheologică*, Ed. Karl A. Romstorfer, Suceava.
- MACFARLAND 2014 MacFarland, T. W., *Introduction to Data Analysis and Graphical Presentation in Biostatistics with R. Statistics in the Large*, Springer, Cham.
- MANTU 1998 Mantu, C.-M., *Cultura Cucuteni. Evoluție, Cronologie, Legături*, Ed. Nona, Piatra-Neamț.
- MARINESCU-BÎLCU 1991 Marinescu-Bîlcu, S., *Sur quelques problemes du néolithique et du énéolithique à l'Est des Carpates Orientales*, in: *Dacia*, N.S., XXXV, 5-60.
- MARINESCU-BÎLCU 1981 Marinescu-Bîlcu, S., *Tîrpești. From Prehistory to History in Eastern Romania*, Archaeopress, Oxford.

- MARINESCU-BÎLCU, BOLOMEY 2000 Marinescu-Bîlcu, S., Bolomey, A., *Drăgușeni. A Cucutenian Community*, Ed. Enciclopedică-Wasmuth Verlag, București-Tübingen.
- MATASĂ 1946 Matasă, C., *Frumușica. Village préhistorique à céramique peinte dans la Moldavie du nord, Roumanie*, Monitorul Oficial și Imprimeriile Statului Imprimeria Națională București.
- MAXIM-KALMAR, TARCEA 1999 Maxim-Kalmar, Z., Tarcea, L., *Analiza matematică și statistică a ceramicii și plasticii civilizației Cucuteni de la Trușești-Țuguieta*, in: Petrescu-Dîmbovița, M., Florescu, M., Florescu, A. C., *Trușești. Monografie arheologică*, Ed. Academiei Române, București – Iași, 647-673.
- MILLER 1986 Miller, R. J., *Beyond ANOVA, Basics of Applied Statistics*, John Wiley & Sons, Inc, New York.
- PECK, OLSEN, DEVORE 2016 Peck, R., Olsen, C., Devore, J. L. *Introduction to Statistics and Data Analysis*, Cengage Learning, Boston.
- PETRESCU-DÎMBOVIȚA 1996 Petrescu-Dîmbovița, M., *Quelques considérations sur la position chronologiques de la station appartenant à l'étape Cucuteni A de Trușești dans le cadre de cette etape*, in: Dumitroaia, Gh., Monah, D. (eds.), *Cucuteni aujourd'hui. 110 ans depuis la découverte en 1884 du site eponyme*, Complexul Muzeal Județean Neamț, Piatra-Neamț, 15-25.
- PETRESCU-DÎMBOVIȚA, FLORESCU, FLORESCU 1999 Petrescu-Dîmbovița, M., Florescu, M., Florescu, A. C., *Trușești. Monografie arheologică*, Ed. Academiei Române, București - Iași.
- R CORE TEAM 2019 R Core Team, *R: A language and environment for statistical computing*, in: The R Project for Statistical Computing, <https://www.R-project.org/> (14.03.2019).
- REVELLE 2018 Revelle, W., *psych: Procedures for Psychological, Psychometric, and Personality Research*, in: The R Project for Statistical Computing, <https://CRAN.R-project.org/package=psych> (14.03.2019).
- RSTUDIO TEAM 2016 RStudio Team, *RStudio: Integrated Development Environment for R*, <http://www.rstudio.com/> (14.03.2019)..
- SARKAR 2008 Sarkar, D., *Lattice: Multivariate Data Visualization with R*, Springer, New York.
- SOROCHIN 2002 Sorochin, V., *Aspectul regional cucutenian Drăgușeni-Jura*, Ed. „Constantin Matasă”, Piatra Neamț.
- TABACHNICK, FIDELL 2013 Tabachnick, B. G., Fidell, L. S., *Using Multivariate Statistics*, Pearson Education Inc., Boston.
- URSULESCU, IGNĂTESCU 2003 Ursulescu, N., Ignătescu, S., *Preutești – „Haltă”. O așezare cucuteniană pe valea Șomuzului Mare*, Casa Editorială Demiurg, Iași.
- VORNICU 2017 Vornicu, D. M., *The chipped stone assemblage from the Early Chalcolithic settlement at Isaiia – Balta Popii*, in: *Materiale și Cercetări Arheologice*, S.N., XIII, 191-211.
- WICKHAM 2016 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag, New York.
- WICKHAM, BRYAN 2019 Wickham, H., Bryan, J., *readxl: Read Excel Files*, in: The R Project for Statistical Computing, <https://CRAN.R-project.org/package=readxl> (14.03.2019).

## LIST OF ILLUSTRATIONS

- Fig. 1. Boxplot presenting the distribution of the values of analysed groups against the median and interquartile range (Dr=Drăgușeni; Pr=Preutești; Fr=Frumușica; Tr=Trușești; Ta=Târpești; Ha=Hăbășești).
- Fig. 2. The distribution by groups of the true value of the mean for a 95% confidence interval.
- Fig. 3. Density distribution by group.
- Fig. 4. ANOVA residuals distribution against fitted values.
- Fig. 5. ANOVA residuals distribution against a hypothetical normal distribution.

## TABLES LIST

- Table 1. Excerpt from the data set.
- Table 2. Descriptive statistics for the data set.
- Table 3. Results of Shapiro-Wilks test for distribution normality.
- Table 4. Results of Levene test for homogeneity of variance.
- Table 5. Analysis of variance results.
- Table 6. Analysis of variance results following natural logarithmic transformation of the data.
- Table 7. Analysis of variance results following scaling of the data.
- Table 8. Scheffe test results.