Aplicații pentru arheologie ale unor concepte de analiză statistică univariată. Teorie și studiu de caz asupra variației formei paharelor din șase așezări ale fazei Cucuteni A ale fazei Cucuteni A

George Bodi

March 1, 2019

# Importul datelor

#se stabilește dosarul de lucru  
setwd("C:/Users/George Bodi/OneDrive/Desktop/Articol ArhMold")   
#se încarcă pachetul necesar citirii documentului excel  
library(readxl)   
#se încarcă documentul excel cu specificarea tipului fiecărei variabile  
pahare <- read\_excel("pahare.xlsx", col\_types = c("text",   
 "numeric"))   
#se transformă documentul excel într-un format prelucrabil cu R  
pahare=as.data.frame(pahare)   
#se cere afișarea aleatorie a 6 rânduri din setul de date  
pahare[sample(nrow(pahare), 6), ]

## Asezare Raport  
## 77 Dr 1.29  
## 88 Dr 0.94  
## 72 Dr 1.03  
## 61 Ha 1.09  
## 3 Ta 1.28  
## 216 Fr 1.17

Tipuri de variabile. În tabelul de mai sus, 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 in 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.

# Curățarea datelor

#definim variabila categorică ca factor, pentru a o putea utiliza în gruparea datelor numerice  
pahare$Asezare=as.factor(pahare$Asezare)

# Analiza exploratorie a datelor

#afișare a datelor sumare a setului de date  
summary(pahare)

## Asezare Raport   
## Dr:32 Min. :0.680   
## Fr:37 1st Qu.:1.030   
## Ha:42 Median :1.085   
## Pr: 4 Mean :1.084   
## Ta:22 3rd Qu.:1.150   
## Tr:95 Max. :1.430

Acest tabel ne prezintă numărul de observații pentru fiecare așezare, precum și valorile minuime, maxime, quartile și centrale pentru întreg setul de date. Media și mediana sunt măsuri ale tendinței centrale. Media (Mean) reprezintă media aritmetică a cărei valoare este obținută după formula Suma/Nr. Mediana (Median) este valoarea care împarte setul de date în două jumătăți egale. Quantilele, ca măsuri ale dispersiei, 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. Diferența dintre valoarea maximă și valoarea minimă definesc amplitudinea unei variabile și oferă o imagine sintetică a întinderii acesteia. Pentru că afișarea statisticilor rezumative pentru întreg setul de date nu ne oferă prea multe informații despre caracteristicile fiecărei așezări, cerem în continuare afișarea acestora pe grupuri. De asemenea, ne dorim și afișarea unor valori suplimentare, pentru a ne ofri o idee mai completă asupra structurii setului de date.

#cerem încărcarea pachetului psych necesar pentru accesarea funcției describeBy  
library("psych", lib.loc="~/R/win-library/3.5")  
##afișare a datelor sumare a setului de date  
describeBy(pahare, group = pahare$Asezare)

##   
## Descriptive statistics by group   
## group: Dr  
## vars n mean sd median trimmed mad min max range skew  
## Asezare\* 1 32 1 0.00 1 1 0.00 1.00 1.00 0.00 NaN  
## Raport 2 32 1 0.12 1 1 0.09 0.68 1.29 0.61 -0.03  
## kurtosis se  
## Asezare\* NaN 0.00  
## Raport 0.51 0.02  
## --------------------------------------------------------   
## group: Fr  
## vars n mean sd median trimmed mad min max range skew  
## Asezare\* 1 37 2.00 0.00 2.00 2.00 0.00 2.00 2.00 0.00 NaN  
## Raport 2 37 1.08 0.08 1.07 1.08 0.06 0.94 1.29 0.35 0.68  
## kurtosis se  
## Asezare\* NaN 0.00  
## Raport 0.27 0.01  
## --------------------------------------------------------   
## group: Ha  
## vars n mean sd median trimmed mad min max range skew  
## Asezare\* 1 42 3.00 0.00 3.00 3.00 0.00 3.00 3.00 0.00 NaN  
## Raport 2 42 1.15 0.09 1.15 1.15 0.09 0.91 1.33 0.42 -0.27  
## kurtosis se  
## Asezare\* NaN 0.00  
## Raport -0.45 0.01  
## --------------------------------------------------------   
## group: Pr  
## vars n mean sd median trimmed mad min max range skew  
## Asezare\* 1 4 4.00 0.00 4.00 4.00 0.00 4.00 4.00 0.00 NaN  
## Raport 2 4 1.07 0.08 1.11 1.07 0.02 0.95 1.13 0.18 -0.7  
## kurtosis se  
## Asezare\* NaN 0.00  
## Raport -1.72 0.04  
## --------------------------------------------------------   
## group: Ta  
## vars n mean sd median trimmed mad min max range skew  
## Asezare\* 1 22 5.0 0.00 5.00 5.0 0.0 5.0 5.00 0.00 NaN  
## Raport 2 22 1.1 0.12 1.08 1.1 0.1 0.8 1.32 0.52 -0.22  
## kurtosis se  
## Asezare\* NaN 0.00  
## Raport -0.05 0.03  
## --------------------------------------------------------   
## group: Tr  
## vars n mean sd median trimmed mad min max range skew  
## Asezare\* 1 95 6.00 0.0 6.00 6.00 0.00 6.00 6.00 0.00 NaN  
## Raport 2 95 1.08 0.1 1.09 1.09 0.09 0.76 1.43 0.67 -0.25  
## kurtosis se  
## Asezare\* NaN 0.00  
## Raport 1.44 0.01

Primul panou al rezultatului ne indică ordinea în care ne sunt prezentate statisticile rezumative pe așezări. Observăm că, în plus față dde primul caz, și de interes pentru noi, ne sunt prezentate măsuri noi ale variației, precum și valori diagnostic pentru caracterizarea distribuției datelor. Măsuri ale variației. O primă măsură a variației suplimentară este constituită de abaterea standard (sd). Această valoare ne indică măsura în care valorile individuale ale unui variabile variază față de medie.  
O a doua măsură oferită este eroarea standard (se). 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. Valorile extreme sunt constituite de puncte specifice cu valori 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. Identificarea valorilor extreme este importantă deoarece pot influența validitatea rezultatelor analizei datelor. De interes sunt și două măsuri ale distribuției care sunt importante în special pentru determinarea tipului de analiză statistică ce urmează a fi utilizat. Indicele de asimetrie ne este oferit în coloana identificată cu skew. În cazul unei distribuții normale, care din punct de vedere grafic poate fi reprezentată ca o curbă gauss, valoarea acestuia este 0. Indicele de aplatizare ne este oferit în coloana identificată cu kurtosis. Într-o distribuție normală valoarea acestuia este 3. Când cei doi indici au valorile de 0 și respectiv 3, seturile de date pot fi analizate utilizând tehnici statistice parametrice. Totuși, valorile de 0 și 3 respectiv sunt 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, 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 mai mari decât 3 ale indicelui de aplatizare 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 valorile distribuite mai larg în jurul mediei, și cu o probabilitate mai mică de apariție a valorilor extreme. În cazul nostru, valorile mici ale indicelui de asimetrie din tabel ne indică o distribuție cu o asimetrie medie, iar valorile indicelui de aplatizare ne indică prezența unor distribuții platykurtice

În continuare vom completa aceste date prin calcularea intervalului de încredere de 95% pentru valorile mediei fiecărui grup, care ne va permite să testăm dacă există eventuale diferențe observabile în estimarea acestora.

#cerem încărcarea pachetului Rmisc necesar pentru accesarea funcției summarySE  
library("Rmisc", lib.loc="~/R/win-library/3.5")

## Loading required package: lattice

## Loading required package: plyr

#cerem calcularea datelor statistice sintetice, cu precizarea setului de date pentru care se face calculul,  
#a variabilei care conține măsurătorile, și a variabilei care conține indicarea așezărilor  
sumar\_pahare=summarySE(pahare, measurevar = "Raport", groupvars = "Asezare")  
#cerem afișarea datelor statistice sintetice  
sumar\_pahare

## Asezare N Raport sd se ci  
## 1 Dr 32 0.996875 0.12169336 0.02151255 0.04387514  
## 2 Fr 37 1.081351 0.07528325 0.01237649 0.02510068  
## 3 Ha 42 1.146905 0.09369093 0.01445682 0.02919615  
## 4 Pr 4 1.075000 0.08426150 0.04213075 0.13407885  
## 5 Ta 22 1.095455 0.12393267 0.02642253 0.05494867  
## 6 Tr 95 1.084526 0.10258124 0.01052461 0.02089687

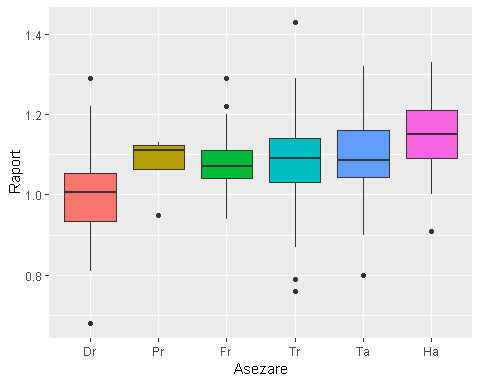
Un prim aspect care se observă din examinarea datelor de mai sus constă în numărul mic de observații (doar patru), care provin din așezare de la Preutești și care sunt caracterizate prin cele mai mari valori ale erorii standard și intervalului de încredere, indicând o variabilitate mare a raportului 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. ## Vizualizarea datelor statistice sintetice și a diferențelor caracteristicilor între așezări

# ordonarea așezărilor în funcție de valoarea medie a raportului  
pahare$Asezare=with(pahare, reorder(Asezare, Raport, mean))   
# cerem încărcarea pachetului ggplot necesar pentru afișarea grafică a datelor statistice sintetice  
library("ggplot2", lib.loc="~/R/win-library/3.5")

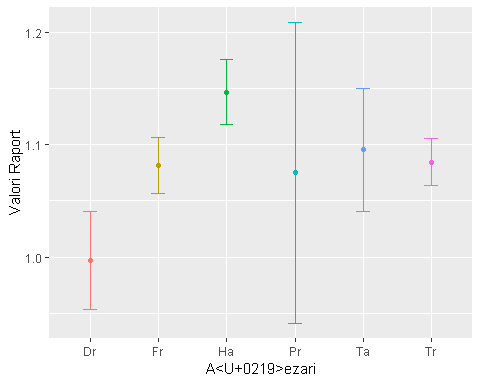
##   
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':  
##   
## %+%, alpha

# cerem generarea unui grafic de tip boxplot cu precizarea setului de date, a variabilei pentru axa x, variabilei pentru axa y și colorarea în funcție de așezare  
ggplot(pahare, aes(x=Asezare, y=Raport, fill=Asezare)) +   
geom\_boxplot() + # precizăm tipul de grafic ca boxplot  
theme(legend.position="none") # eliminăm afișarea legendei graficului

 Diagrama boxplot 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.

# cerem generarea unui grafic pentru afișarea intervalelor de eroare a mediei cu un interval de încredere de 95% cu precizarea setului de date, a variabilei pentru axa x, variabilei pentru axa y și colorarea în funcție de așezare  
ggplot(sumar\_pahare, aes(x=sumar\_pahare$Asezare, y=sumar\_pahare$Raport, colour = sumar\_pahare$Asezare))+  
   
geom\_errorbar(aes(ymin=Raport-ci, ymax=Raport+ci), width=.2)+ # precizăm tipul de grafic ca bare de eroare cu calcularea intervalului de încredere față de medie   
geom\_point()+ # cerem afișarea mediei ca punct pe bara de eroare  
labs(x="Așezări", y="Valori Raport")+ # specificăm titlurile axelor x și Y  
theme(legend.position="none") # eliminăm afișarea legendei graficului

 Reprezentarea grafică a intervalelor de încredere de 95%, ne ajută să vizualizăm probabilitățile de situare 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 valoare 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.

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 analizarea 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.

## Calcularea normalității distribuției

# cerem încărcarea pachetului ggplot necesar pentru afișarea grafică a datelor statistice sintetice  
library("MVN", lib.loc="~/R/win-library/3.5")

## sROC 0.1-2 loaded

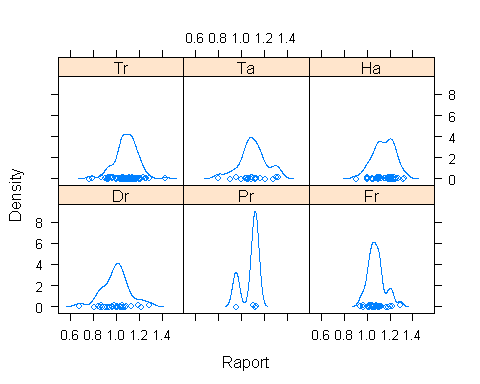
# cerem verificarea normalității univariate a datelor cu precizarea setului de date și a grupării măsurătorilor pe așezări  
mvnpahare=mvn(data=pahare, subset = "Asezare")  
# cerem afișarea rezultatelor testului de analiză univariată a normalității  
mvnpahare$univariateNormality

## $Dr  
## Test Variable Statistic p value Normality  
## W Shapiro-Wilk variable 0.9786801 0.7603894 YES  
##   
## $Pr  
## Test Variable Statistic p value Normality  
## W Shapiro-Wilk variable 0.7576391 0.04541828 NO  
##   
## $Fr  
## Test Variable Statistic p value Normality  
## W Shapiro-Wilk variable 0.954128 0.1310867 YES  
##   
## $Tr  
## Test Variable Statistic p value Normality  
## W Shapiro-Wilk variable 0.9723423 0.04162913 NO  
##   
## $Ta  
## Test Variable Statistic p value Normality  
## W Shapiro-Wilk variable 0.968886 0.6852827 YES  
##   
## $Ha  
## Test Variable Statistic p value Normality  
## W Shapiro-Wilk variable 0.9794636 0.6408369 YES

Primul panou al rezultatului ne indică ordinea în care ne sunt prezentate rezultatele testării normalității distribuției datelor în funcție de așezări. Observăm că majoritatea datelor sunt considerate ca prezentând o distribuție normală, cu excepția așezărilor de la Preutești și Trușești.Pentru siguranță vom efectua și o inspecție vizuală a distribuției datelor cu ajutorul graficelor de densitate.

## Vizualizarea normalității distribuției

# cerem încărcarea pachetului ggplot necesar pentru afișarea grafică a distribuției datelor  
library("lattice", lib.loc="C:/Program Files/Microsoft/R Open/R-3.5.1/library")  
# cerem generarea unui grafic de distribuție a densității datelor, cu precizarea proiectării măsurătorilor pe variabila așezărilor și cu specificare setului de date  
densityplot(~ Raport | Asezare, data = pahare)

 Deși testul Shapiro-Wilks respinge ipoteza unei distribuții normale în cazul așezării de la Trușești, noi o vom accepta în baza inspecției vizuale a graficului densității și, datorită numărului mare de observații, în baza teoremei de limită centrală. 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 4 măsurători.

#eliminăm din analiză valorile provenite din așezarea de la Preutești  
pahare2=pahare[-which(pahare$Asezare=="Pr"), ]

Cea de a doua condiție a validității unei analize ANOVA constă în existența unei variații comune în interiorul grupurilor.

# Analiza variației comunune

# cerem încărcarea pachetului car necesar accesării funcției leveneTest  
library(car)

## Loading required package: carData

##   
## Attaching package: 'car'

## The following object is masked from 'package:psych':  
##   
## logit

# efectuăm testul Levene precizând ca dorim analizarea măsurătorilor pe grupuri în funcție de așezare  
leveneTest(pahare2$Raport~pahare2$Asezare)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)  
## group 4 1.3581 0.2495  
## 223

Rezultatele testului Levene cu valoarea p>0.05 ne permite să acceptăm ipoteza conform căreia variația între cele cinci așezări rămase în analiză este omogenă.

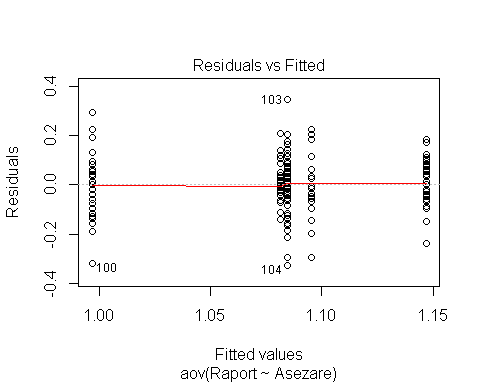
### Analiza variației (ANOVA) și verficarea distribuției valorilor reziduale

#efectuăm analiza variației măsurătorilor în funcție de grupuri  
anv\_pahare=aov(Raport ~ Asezare, data = pahare2)  
#cerem afișarea rezultatului analizei  
summary.aov(anv\_pahare)

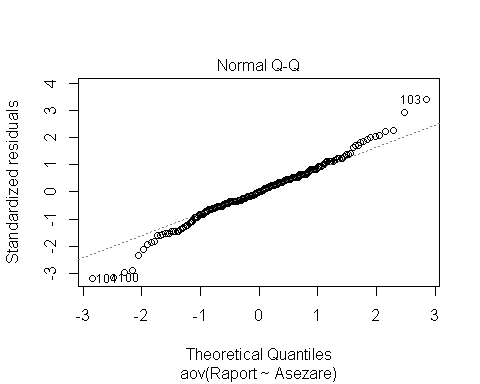
## Df Sum Sq Mean Sq F value Pr(>F)   
## Asezare 4 0.4123 0.10306 9.844 2.37e-07 \*\*\*  
## Residuals 223 2.3347 0.01047   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Valoarea p, puternic negativă ne indică faptul că există diferențe semnificative statistic între dimensiunile paharelor specifice fiecărei așezări, constatare confirmată și de valoare F. Urmează să verificăm validitatea analizei prin inspecția distribuției valorilor reziduale.

plot(anv\_pahare, 1)

 Diagrama de distribuție a rezidualelor față de linia valorilor medii, unde nu se observă o relație evidentă între acestea, ne confirmă faptul că există omogenitate a variației.

plot(anv\_pahare, 2)

 Acordul dintre distribuția rezidualelor față de linia ipotetică a unei distribuții normale ne indică, de asemenea, faptul că acestea urmează o distribuție normală și că rezultatele noastre sunt valide.

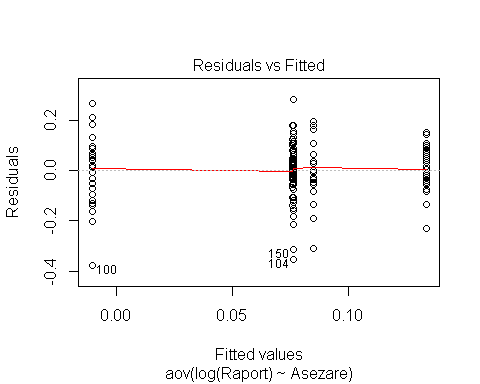
Cele două grafice a distribuției rezidualelor ne semnalează, de asemenea, prezența a trei valori extreme situate la rândurile 100, 103 și 104 în setul nostru de date. Deoarece nu putem să le atribuim unor erori de colectare a datelor, acceptăm faptul că aceste trei cazuri nu pot fi eliminate din analiză.

Putem încerca să le micșorăm impactul asupra analizei noastre prin transformarea logaritmică a datelor în bază naturală sau prin centrarea acestora în urma scăderii valorii mediei din fiecare măsurătoare și împărțirea rezultatului la valoarea deviației standard.

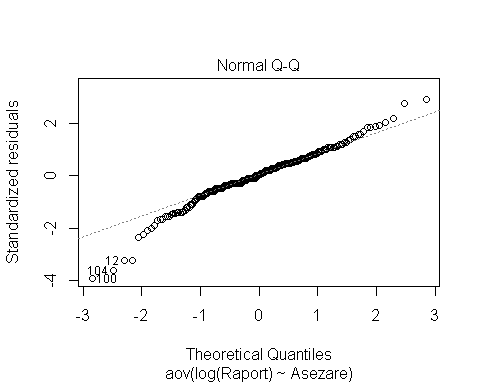
# transformare logaritmică în bază naturală  
anv\_pahare2=aov(log(Raport) ~ Asezare, data = pahare2) #funcția log transformă valorile raportului în bază naturală  
summary.aov(anv\_pahare2)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Asezare 4 0.3815 0.09538 10.05 1.71e-07 \*\*\*  
## Residuals 223 2.1173 0.00949   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

plot(anv\_pahare2, 1)



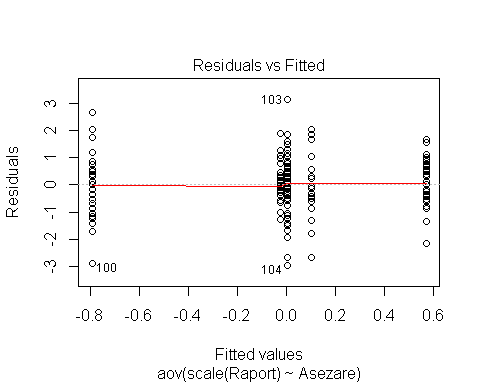
plot(anv\_pahare2, 2)



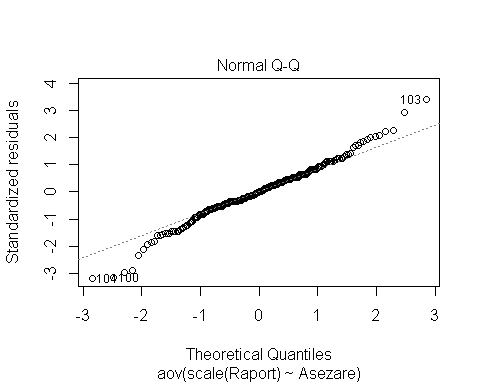
# centrarea datelor  
anv\_pahare3=aov(scale(Raport) ~ Asezare, data = pahare2) #funcția scale centrează valorile  
summary.aov(anv\_pahare3)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Asezare 4 34.07 8.517 9.844 2.37e-07 \*\*\*  
## Residuals 223 192.93 0.865   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

plot(anv\_pahare3, 1)



plot(anv\_pahare3, 2)



Observăm că rezultatele nu sunt cu mult îmbunătățite, de unde deducem că impactul acestor valori nu este semnificativ, astfel încât ne continuăm analiza pe setul de date netransformate.

### Testul Scheffe pentru analiză a variației pe perechi de grupuri

#cerem încărcarea pachetului agricolae pentru accesarea funcției scheffe.test  
library("agricolae", lib.loc="~/R/win-library/3.5")  
#efectuăm testul scheffe pe modelul generat de analiza anova cu specificarea faptului că dorim afișarea rezultatelor acestuia și că utilizăm date grupate  
scheffe.test(anv\_pahare, "Asezare", console = TRUE, group = TRUE)

##   
## Study: anv\_pahare ~ "Asezare"  
##   
## Scheffe Test for Raport   
##   
## Mean Square Error : 0.01046958   
##   
## Asezare, means  
##   
## Raport std r Min Max  
## Dr 0.996875 0.12169336 32 0.68 1.29  
## Fr 1.081351 0.07528325 37 0.94 1.29  
## Ha 1.146905 0.09369093 42 0.91 1.33  
## Ta 1.095455 0.12393267 22 0.80 1.32  
## Tr 1.084526 0.10258124 95 0.76 1.43  
##   
## Alpha: 0.05 ; DF Error: 223   
## Critical Value of F: 2.412129   
##   
## Groups according to probability of means differences and alpha level( 0.05 )  
##   
## Means with the same letter are not significantly different.  
##   
## Raport groups  
## Ha 1.146905 a  
## Ta 1.095455 ab  
## Tr 1.084526 b  
## Fr 1.081351 b  
## Dr 0.996875 c

Testul post hoc pentru analiza variației pe perechi de grupuri, prezentat în panourile de mai sus, ne confirmă imaginea construită de vizualizarea intervalelor de încredere. În acest moment putem afirma cu încredere 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 standard = 0.1) și Frumușica (media = 1.08, abatere standard = 0.07) sunt specifice de asemenea de 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 de0.1. 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.