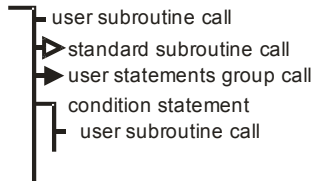


Schematy

Legend
'comment

subroutine name



iteration loop with range values

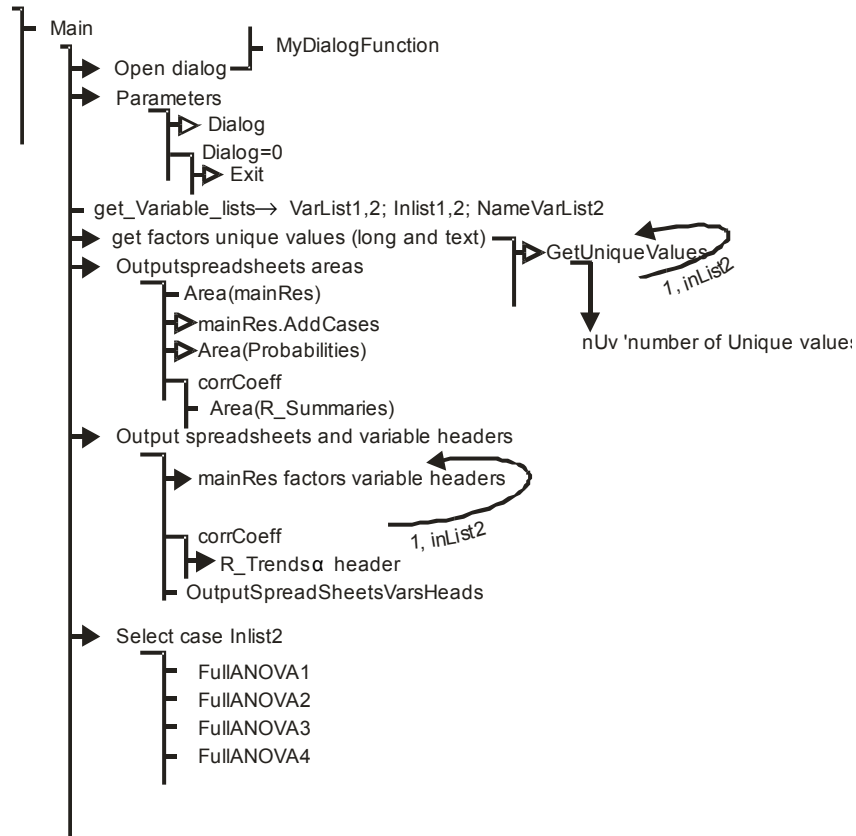


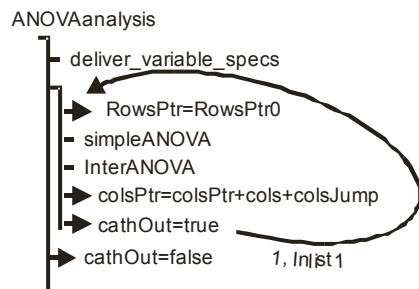
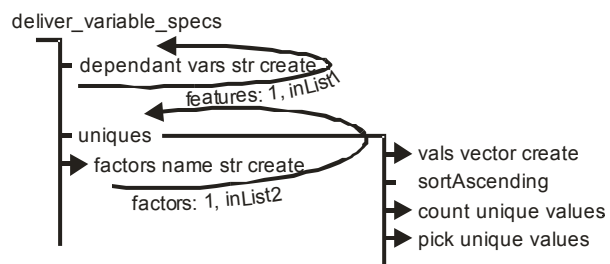
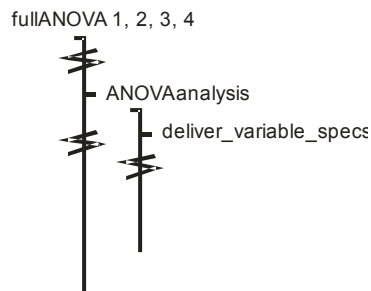
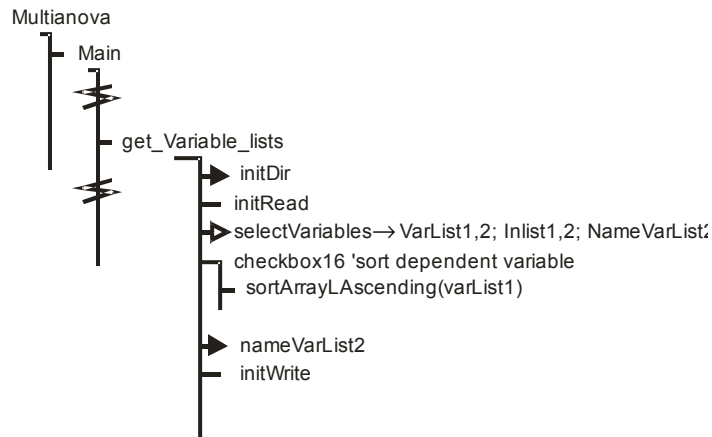
subroutine name



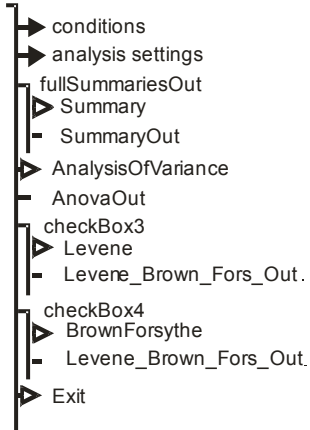
Statistica 10
Read data
Multanova

Multanova

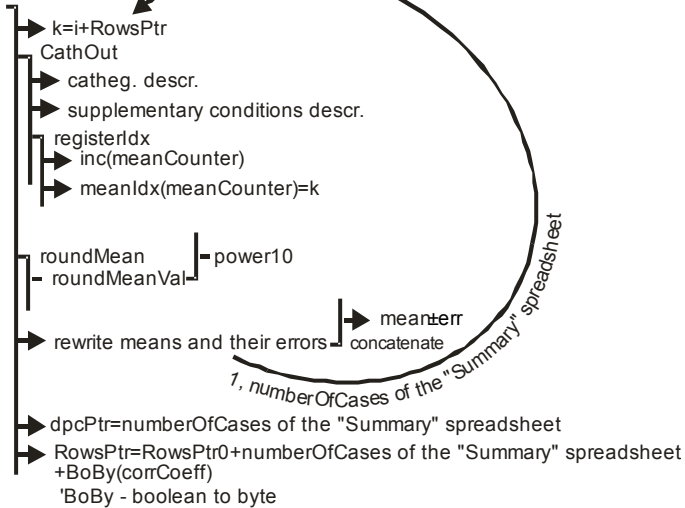




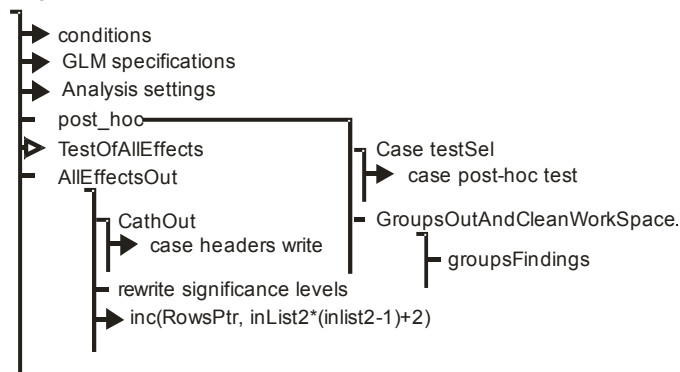
SimpleANOVA



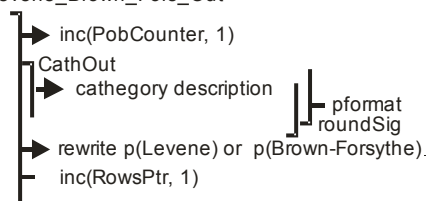
SummaryOut



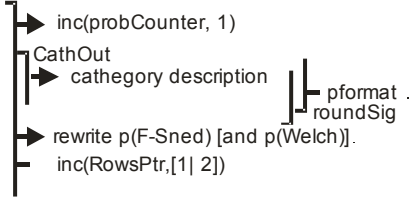
InterANOVA



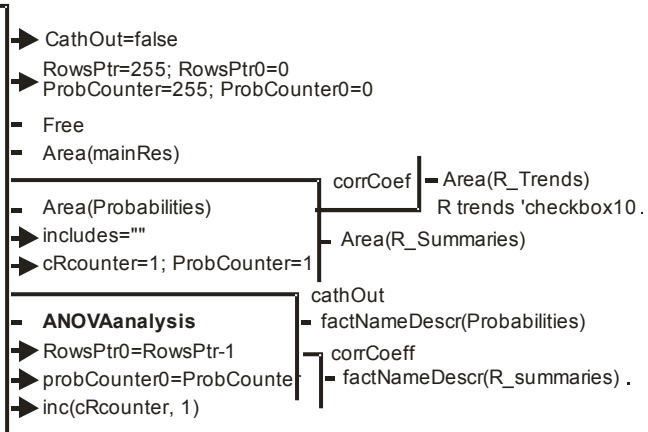
Levene_Brown_Fors_Out



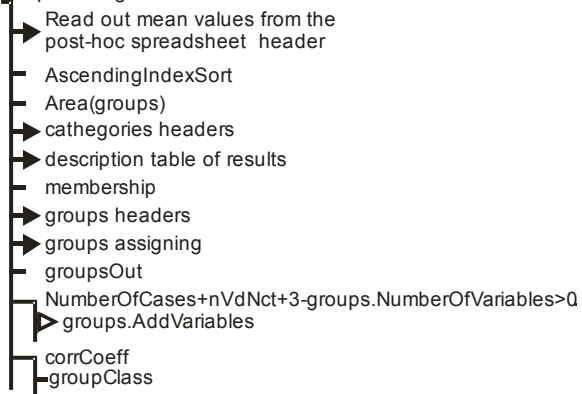
AnovaOut



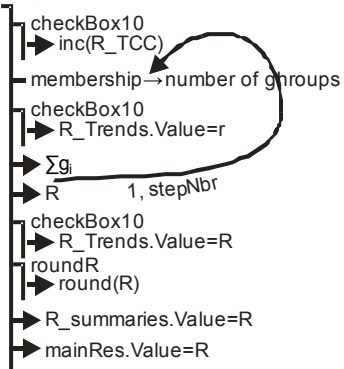
FullANOVA1



GroupsFindings



groupClass



'checkBox10 - create R_Trends spreadsheet'

FullANOVA2

```

➔ CathOut=false
➔ RowsPtr=255; RowsPtr0=0
➔ ProbCounter=255; ProbCounter0=0
- Free
- Area(mainRes)
- Area(Probabilities)
➔ VarList_21(1)=VarList2(InList2)
➔ cRcounter=1; ProbCounter=1
- includes="("+namevarlist2(InList2-1)+"="+" Str(valuesArr(InList2-1,k1))+")"
- AnovaAnalysis
➔ RowsPtr0=RowsPtr-1
➔ ProbCounter0=ProbCounter
➔ inc(cRcounter, 1)
  1, nUv(inlist2-1)
- cathOut
- factNameDescr(Probabilities)
- corrCoef
- factNameDescr(R_summaries) .

➔ RowsPtr0=rowsPtr
➔ ProbCounter0=ProbCounter
➔ VarList_22 write down
➔ includes=""
- AnovaAnalysis

```

FullANOVA3

```

➔ CathOut=false
➔ RowsPtr=255; RowsPtr0=0
➔ ProbCounter=255; ProbCounter0=0
- Free
- Area(mainRes)
➔ Area(Probabilitiess)
➔ VarList_21(1)=VarList2(InList2)
- CathDescr
➔ t_includes="("+ namevarlist2(InList2-2)+"="+ Str(valuesArr(InList2-2,k2))+")"
- CathDescr
➔ includes=t_includes+" and ("+"+namevarlist2(InList2-1)+"="+ Str(valuesArr(InList2-1,k1))
- ANOVAAnalysis
➔ RowsPtr0=RowsPtr-1
➔ cRcounter=cRcounter+1
➔ CathOut 1, nUv(inlist2-1)
➔ cathegory description
➔ VarList_22 write down
➔ RowsPtr0=rowsPtr
➔ ProbCounter0=ProbCounter
➔ includes=t_includes
- AnovaAnalysis
➔ RowsPtr0=rowsPtr-1
➔ ProbCounter0=ProbCounter
➔ inc(cRcounter)
➔ CathOut 1, nUv(itis2-2)
➔ cathegory description
➔ RowsPtr0=rowsPtr
➔ ProbCounter0=ProbCounter
➔ includes=""
- AnovaAnalysis

```

corrCoef | Area(R_Trends)
R trends 'checkbox10.
- Area(R_Summaries)

FullANOVA4

```

→ cathOut=False; RowsPtr=255; RowsPtr0=0; cRcounter=1;
Free
Area(mainRes)
Area(Probabilities)
VarList_21 write down
CathDescr
t2_includes="(" + namevarlist2(InList2-3) + "" = " + Str(valuesArr(InList2-3,k3)) + ")"
CathDescr
t1_includes=t2_includes + " and (" + namevarlist2(InList2-2) + "" = " + Str(valuesArr(InList2-2,k2)) + ")"
CathDescr
includes=t1_includes + " and (" + namevarlist2(InList2-1) + "" = " + Str(valuesArr(InList2-1,k1)) + ")"
ANOVAAnalysis
RowsPtr0=RowsPtr-1
ProbCounter0=ProbCounter
inc(cRcounter, 1)
CathOut
    cathegory description
VarList_22 write down
RowsPtr0=rowsPtr
ProbCounter0=ProbCounter
includes=t1_includes
AnovaAnalysis
RowsPtr0=rowsPtr-1
ProbCounter0=ProbCounter
inc(cRcounter)
CathOut
    cathegory description
VarList_23 write down
RowsPtr0=rowsPtr
ProbCounter0=ProbCounter
includes=t2_includes
AnovaAnalysis
RowsPtr0=rowsPtr-1
ProbCounter0=ProbCounter
inc(cRcounter)
CathOut
    cathegory description
RowsPtr0=rowsPtr
ProbCounter0=ProbCounter
includes=""
AnovaAnalysis

```

corrCoef Area(R_Trends)
R trends 'checkbox10.
Area(R_Summaries)

1, nUv(InList2-2)

1, nUv(InList2-3)

Cele ANOV'y

Grupowanie wynika z istotności różnic pomiędzy średnimi. Zakładane grupowanie osobników wynika z kontrolowanych (np. sposób pielęgnacji) lub znanych wartości czynników (np. pochodzenie).

Zadaniem ANOV-y jest potwierdzenie lub zaprzeczenie słuszności takiego grupowania na podstawie pomiarów pewnej cechy. Wyniki pomiarów będą się na ogół różnić, a różnice będą wynikać nie tylko z przynależności do różnych klas, ale także z istnienia czynników nieuwzględnionych (o których zakłada się, że ich wpływy się niwelują, jeśli bierzemy pod uwagę średnią z pomiarów danej cechy u kilku osobników). Dla badacza interesujące jest, czy i w jakim stopniu czynniki kontrolowane wpływają na interesujące go cechy pewnej populacji. Istotność różnic średnich ocenia się za pomocą statystyki F-Snedecora. Ta, jako ocena punktowa, informuje jednak tylko o występowaniu co najmniej jednej istotnej różnicy pomiędzy danymi klasami, nie wskazując jednak pomiędzy którymi. Po ujawnieniu takiej różnicy potrzebne są dalsze badania oparte na testach post-hoc, które pozwalają na wydzielenie tzw. grup jednorodnych, czyli grup osobników nie różniących się istotnie w wymiarze pewnej cechy.

Liczba grup jednorodnych może się wahać w granicach od 1 (osobniki nie są odróżnialne w wymiarze danej cechy) do liczby klas a priori (klasyfikacja wg wartości wziętych pod uwagę czynników w pełni potwierdza się w wymiarze danej cechy).

Liczba grup zależy od zakładanej a priori progowej wartości poziomu istotności, przy którym orzekamy, że dane dwie średnie nie różnią się istotnie. Podział na grupy jednorodne jest więc obciążony czynnikiem subiektywnym.

Interesujące byłoby znalezienie współczynnika oceniającego zróżnicowanie grup niezależnie od owego progu. Mamy już co prawda taki parametr – jest nim poziom istotności wg statystyki F-Snedecora, ale parametr ten jest niewygodny w użyciu ze względu gwałtowne zmiany w funkcji różnicy pomiędzy średnimi. Najlepszym współczynnikiem byłby taki, który liniowo zależałby od błędu standardowego średnich (ANOVA wymaga jego równości dla wszystkich średnich). Uznaliśmy, że podstawą takiego współczynnika może być suma liczby znalezionych klas po wszystkich wartościach wartości progowej. Aby uniezależnić się od liczby a priori wyróżnianych klas, wartości tej sumy rzutujemy na przedział $\langle 0, 1 \rangle$, czyli normalizujemy. Możliwość ta wynika stąd, iż znane są wartości graniczne tej sumy. Właściwości tego współczynnika wymagają jeszcze badań.

Grupy jednorodne

W tradycyjnym postępowaniu, na podstawie analizy macierzy istotności różnic testu post-hoc (MIR), tworzy się tablicę postaci (rys. 1)¹

średnia	Grupa		
	A	B	C
a1	*		
a6	*	*	
a3		*	
a2		*	
a5		*	*
a4			*

grupy jednorodne:

¹ por. np. http://www.mp.pl/artykuly/index.php?aid=10851&_tc=54B3BE61006FCE959BD864EB246587A3

$A=\{a1,a6\}$
 $B=\{a6, a3, a2, a5\}$
 $C=\{a5, a4\}$

Rys 1. Zasada tworzenia grup jednorodnych

Pierwsza kolumna zawiera posortowane rosnąco średnie, a w następnych, za pomocą *, oznacza się wartości nie różniące się istotnie (przy zadanym poziomie istotności α). W rezultacie tworzy się tzw. „jednorodne” grupy średnich. W powyższym przykładzie będą to grupy A, B, C o podanych elementach. W zestawieniach wyników analizy wariancji poszczególnym średnim przypisuje się symbole grup, do których one należą (rys. 2)

średnia	Grupa
a1	A
a2	B
a3	B
a4	C
a5	BC
a6	AB

Rys. 2. Wykaz średnich z oznaczeniem przynależności do grup jednorodnych

Celem naszego programu jest m.in. automatyczne tworzenie takich zestawień.

Standardowa procedura grupowania średnich zawarta w pakiecie Statistica daje błędne wyniki, dlatego zdecydowaliśmy się na napisanie własnej². Nasza procedura idzie dalej niż standardowa. Ta pozwala tylko odczytać symboliczne przypisanie średnich do grup jednorodnych na podstawie wykresu gwiazdkowego. Nasza procedura dokonuje jawnego przypisania symboli grup w końcowym zestawieniu wyników analiz wariancji.

Zaliczanie średnich do grup jednorodnych odbywa się na podstawie MIR dla wybranego testu post-hoc. W naszym programie wykonuje to procedura „*membership*” wołana z procedury „*groupsFindings*”.

Współczynnik korelacji

Dwie średnie zaliczamy do jednej grupy jednorodnej, gdy hipotezę o ich równości przyjmujemy na poziomie ufności $p > 1 - \alpha$. Im wyższy poziom α , tym więcej grup jednorodnych. W krańcowym przypadku, gdy $\alpha = 1$, każda wartość stanowi grupę jednorodną – liczba grup jednorodnych będzie wtedy równa mocy zbioru wartości średnich.

Wartość α jest wyborem subiektywnym. Aby zbadać jego wpływ na wynik grupowania przeglądamy grupowanie dla każdego α . Waga subiektywności wyboru tej wartości jest tym mniejsza, im grupowanie będzie mniej od tego wyboru zależeć. Grupowanie może być niezależne od poziomu α w przypadku, gdy błędy standardowe średnich są równe 0, bądź w przypadku, gdy wszystkie wartości badanej cechy są sobie równe. Dla oceny przydatności danej cechy dla rozróżniania elementów populacji może być użyta liczba klas wyróżnianych przez jej wartości. Funkcja do tego celu powinna być:

- 1) proporcjonalna do liczby wyróżnianych klas,
- 2) odwrotnie proporcjonalna do liczby a priori wyróżnianych klas,
- 3) powinna przyjmować wartości z przedziału $\langle 0,1 \rangle$.

² Nie dotyczy to pakietu Statistica 5.

Poszukiwana funkcja wartość 0 powinna przyjmować, gdy wszystkie elementy populacji zostały zaliczone do jednej grupy, a 1 – gdy liczba grup jest równa liczbie klas wyróżnionych a priori. Warunek 1. i 2. spełnia iloraz

$$r = \frac{g}{c},$$

gdzie g , to liczba grup jednorodnych, a c , to liczba klas. Informację, jak te warunki są spełniane w zależności od α , zawiera funkcja

$$J = \int_0^1 r(\alpha) d\alpha.$$

Ponieważ $1 \leq g \leq c$, więc $J \in \left\langle \frac{1}{c}, 1 \right\rangle$. Przedział ten należy rzutować na przedział $\langle 0, 1 \rangle$ za pomocą transformacji $y = ax + b$. Rozwiązanie układu równań

$$\begin{cases} \frac{a}{c} + b = 0 \\ a + b = 1 \end{cases}$$

daje $a = \frac{c}{c-1}$ oraz $b = -\frac{1}{c-1}$ i w konsekwencji, oznaczając postać znormalizowaną symbolem R , mamy

$$R = \frac{1}{c-1} (J - 1)$$

W obliczeniach całkę oznaczoną zastępujemy sumowaniem, i przy założeniu że $r(\alpha)$ próbujemy w równych odstępach, dochodzimy do formuły

$$R = \frac{1}{c-1} \left(\frac{1}{n} \sum_{i=1}^n g_i - 1 \right),$$

gdzie n , to liczba punktów, w wymiarze poziomu istotności α , g_i – liczba grup jednorodnych dla danego zbioru średnich wyznaczonych przy wartości progowej $\alpha = \frac{i}{n}$, dla $i = 1 \dots n$. Współczynnik ten wyraża zależność badanej cechy od czynników zewnętrznych, czyli zmiennej zależnej od zmiennych niezależnych, dlatego nazywamy go współczynnikiem korelacji. W programie naszym potrzebne obliczenia wykonuje procedura *groupClass*.

groupClass. Na podstawie MIR procedura ta wyznacza grupy jednorodne średnich w aktualnie rozpatrywanej klasyfikacji dla zadanej liczby n poziomów α . Po każdym grupowaniu obliczany jest współczynnik R i wyprowadzany do głównego arkusza wyników pod kolumną wykazu przynależności średnich do grup jednorodnych. Program wyprowadza także wyniki pośrednie $r(i)$ do arkusza *R_Trends* oraz zestawienie wszystkich R do arkusza *R_summaries*. Procedurę rozpoczyna określenie kolumny w arkuszu *mainRes*, do której ma być wyprowadzona wartość R , powiększany jest licznik wierszy *R_TCC* arkusza *R_Trends* i zapisywany jest nagłówek kolumny wyników $R(\alpha)$. Nagłówek ten zawiera nazwę zmiennej odczytaną z arkusza danych oraz wartości klasyfikatorów. Następnie, w pętli iteracyjnej, dla kolejnych wartości progowych α wołana jest procedura *membership*, która dostarcza informacji o liczbie grup jednorodnych znalezionych przy danej wartości progowej α . Liczba ta jest podstawą do obliczeń współczynnika R – obliczany jest iloraz r oraz suma wartości g . Obliczane kolej-

no wartości $r_i = \frac{g_i}{c}$ są zapisywane do arkusza *R_Trends*. Po wyjściu z pętli iteracyjnej obliczany jest współczynnik R i zapisywany do arkusza *R_Trends*, *R_summaries* i do *mainRes*.

Tabela 6. Współrzędne zapisów wartości R do arkuszy wyników

arkusz	wiersz	kolumna
R_Trends	$alphaStepNbr+1$	R_TCC+1
$R_summaries$	$cRcounter$	$offset+InList2+varNbr$
$mainRes$	$meanIdx(meanCounter)+1$	$InList2+ColsPtr+cols+1+offset+groupshift$

$alphaStepNb = n$ – liczba punktów α , R_TCC – licznik kolumn w arkuszu R_Trends

$cRcounter$ – licznik wierszy w arkuszu $R_summaries$, $inlist2$ – liczba czynników, $offset=inlist2$ – przesunięcie kolumny, $varNbr$ – numer zmiennej zależnej

$meanId$ – skrowidz położenia średnich w arkuszu $mainRes$, $meanCounter$ – licznik średnich, $ColsPtr$ – wskaźnik wolnej kolumny w arkuszu $mainRes$, $cols$ – przesunięcie kolumn uwzględniające konkatenację \pm , $groupshift$ – rozmiar grupy kolumn

Obliczenia R pochłaniają dużo czasu, co zależy od gęstości próbkowania $r(\alpha)$. Rekomendujemy $n=100$.

groupsFindings. Procedura operuje na macierzy prawdopodobieństw testu post-hoc, MIR (tabela 1). Wyniki działania zapisuje początkowo do pomocniczego arkusza *groups* oraz, za pośrednictwem procedury *groupsOut*, do głównego arkusza wyników „mainResults” i do dwóch dalszych arkuszy zawierających zestawienie współczynników korelacji i przebiegi wartości współczynników korelacji w funkcji poziomu istotności.

Procedura najpierw odczytuje wartości średnich z nagłówka MIR, po czym sortuje je rosnąco dostarczając w wyniku indeksu sortującego. Wynik zapisywany jest do arkusza „groups”. Do arkusza tego zapisuje też indeks sortujący (*IdxVect*) jako „unsorted index”. Wykorzystywany jest on do przeglądania macierzy prawdopodobieństw testu post-hoc w taki sposób, jak gdyby była ona posortowana wg wartości średnich opisujących jej wiersze i kolumny. Stwierdzono też, że porządek tych wartości jest taki sam, jak porządek średnich wyprowadzanych przez standardową procedurę *SimpleAnova*, co wykorzystywane jest później do przypisywania owym średnim symboli grup jednorodnych. Do arkusza tego, dla celów kontrolnych, wyprowadzana jest też klasyfikacja kategorialna średnich – tabela 2. Następnie wołana jest procedura *membership*, która wypełnia kolumny pomocniczej tabeli *groups* wartościami poziomów istotności różnic w grupach jednorodnych. Procedura standardowa wpisuje znaki *. Po wykonaniu procedury *membership* usuwane są zbędne kolumny tabeli *groups*, dodawane są nagłówki kolumn grup jednorodnych, a poszczególnym średnim przypisywane są ciągi symboli oznakowujących grupy jednorodne, do których one należą. Przynależność danej średniej do grup jednorodnych jest odczytywana z wierszy tabeli *groups* – są to te grupy, gdzie w wierszu, na początku którego jest wpisana ta średnia jest wpisana wartość $>\alpha$, a ponieważ tylko takie wartości były wpisywane, to bada się, czy jest to wartość niepusta – u nas będzie to wartość >0 . W kolejnym kroku informacja o przynależności do grup jednorodnych jest zapisywana do zestawienia wynikowego – do arkusza *mainRes*. Problemem jest tu wskazanie odpowiedniego wiersza tego arkusza, zważywszy zwłaszcza, że lokalizacje średnich zależą od wielu opcji. Problem ten rozwiązano w ten sposób, że równoległe z zapisem średnich do arkusza wynikowego, tworzony jest wektor *meanIdx* wskazujący wiersze, w których średnie zostały zapisane. Średnie te jednak są zapisane w kolejności wynikającej z hierarchii klasyfikacyjnej, a w arkuszu grupowania w kolejności rosnącej. W procedurze grupowania odnotowywany jest wektor sortujący *IdxVect* zapisywany w 1. kolumnie arkusza *groups*. Korzystając z obydwu tych wektorów można wskazać wiersz, w którym dana średnia została zapisana w arkuszu wyników, jest to

$$meanIdx(IdxVect(i))$$

gdzie i wskazuje wiersz w arkuszu grupowania. W programie korzystano z zapisu wektora sortującego do arkusza grupowania, więc zamiast *IdxVect* mamy w instrukcji *actDat.Value(i,1)*. Powyższe wyrażenie jest ważne tylko dla pierwszego bloku wyników, takiego jaki wystąpił w arkuszu grupowania. Dla opisu kolejnych takich bloków trzeba uwzględnić też przesunięcie wynikające z wydrukowanych już bloków, stąd w instrukcji obliczeniowej jest

$$meanIdx(actDat.Value(i,1)+j),$$

gdzie $j = meanCounter - actDat.NumberOfCases$, *meanCounter* – bieżąca liczba średnich (w aktualnie podsumowywanym bloku łącznie z już podsumowanymi blokami), *actDat.NumberOfCases* – liczba średnich w bloku.

Tabela 1. Przykładowa macierz prawdopodobieństw testu post-hoc (MIR)

Test NIR; zmienna Fe [ppm] (A._BCF_BAF_TC_2009.sta) Prawdopodobieństwa dla testów post-hoc Błąd: MS międzygrupowe = 1031E3, df = 312,00

	Wariant	Gatunek	Pochodz	{1} – 784,45	{2} – 178,38	{3} – 873,59	{4} – 1268,1	{5} – 336,14	{6} – 1001,0	{7} – 650,20	{8} – 626,76	{9} – 532,31	{10} – 771,93	{11} – 550,35	{12} – 712,89
1	kontrola	A. incana	Krzyż		0,029043	0,747275	0,081116	0,105768	0,433815	0,627448	0,568665	0,362275	0,963869	0,397600	0,795849
2	kontrola	A. incana	Rzepin	0,029043		0,012387	0,000099	0,568533	0,003141	0,088774	0,105726	0,201266	0,032508	0,179295	0,054007
3	kontrola	A. incana	Wolsztyn	0,747275	0,012387		0,154466	0,052706	0,644995	0,419521	0,372469	0,217799	0,713231	0,243055	0,561348
4	kontrola	A. glutinosa	Krzyż	0,081116	0,000099	0,154466		0,000840	0,334682	0,026077	0,020956	0,008163	0,073582	0,009850	0,045413
5	kontrola	A. glutinosa	Rzepin	0,105768	0,568533	0,052706	0,000840		0,016714	0,256658	0,293802	0,478336	0,115837	0,438848	0,173781
6	kontrola	A. glutinosa	Wolsztyn	0,433815	0,003141	0,644995	0,334682	0,016714		0,205210	0,176615	0,090870	0,407723	0,103948	0,297923
7	KGHM	A. incana	Krzyż	0,627448	0,088774	0,419521	0,026077	0,256658	0,205210		0,932458	0,669985	0,659903	0,718138	0,820682
8	KGHM	A. incana	Rzepin	0,568665	0,105726	0,372469	0,020956	0,293802	0,176615	0,932458		0,732767	0,599755	0,782382	0,755493
9	KGHM	A. incana	Wolsztyn	0,362275	0,201266	0,217799	0,008163	0,478336	0,090870	0,669985	0,732767		0,386586	0,947979	0,513960
10	KGHM	A. glutinosa	Krzyż	0,963869	0,032508	0,713231	0,073582	0,115837	0,407723	0,659903	0,599755	0,386586		0,423311	0,830991
11	KGHM	A. glutinosa	Rzepin	0,397600	0,179295	0,243055	0,009850	0,438848	0,103948	0,718138	0,782382	0,947979	0,423311		0,556868
12	KGHM	A. glutinosa	Wolsztyn	0,795849	0,054007	0,561348	0,045413	0,173781	0,297923	0,820682	0,755493	0,513960	0,830991	0,556868	

Tabela 2. Przykładowa tabela grupowania średnich po zapisie średnich w porządku rosnącym

E:\GLP\Inne\A.- K. Ufnalski\A._2009\A._BCF_BAF_TC_2009.sta Test NIR; zmienna Fe [ppm] (A._BCF_BAF_TC_2009.sta); subroutine 'groupsFindings'

	unsorted index	Wariant	Gatunek	Pochodz	sorted means	Zmn6	Zmn7	Zmn8	Zmn9	Zmn10	1	2	3	4	5	6	7	8
1	2	kontrola	A. incana	Krzyż	178,38													
2	5	kontrola	A. incana	Rzepin	336,14													
3	9	kontrola	A. incana	Wolsztyn	532,31													
4	11	kontrola	A. glutinosa	Krzyż	550,35													
5	8	kontrola	A. glutinosa	Rzepin	626,76													
6	7	kontrola	A. glutinosa	Wolsztyn	650,2													
7	12	KGHM	A. incana	Krzyż	712,89													
8	10	KGHM	A. incana	Rzepin	771,93													
9	1	KGHM	A. incana	Wolsztyn	784,45													
10	3	KGHM	A. glutinosa	Krzyż	873,59													
11	6	KGHM	A. glutinosa	Rzepin	1001													
12	4	KGHM	A. glutinosa	Wolsztyn	1268,1													

Tabela powyższa po zapisie 1. grupy jednorodnej

E:\GLP\Inne\ALNUS- K. Ufnalski\Alnus_2009\Alnus_BCF_BAF_TC_2009.sta Test NIR; zmienna Fe [ppm] (Alnus_BCF_BAF_TC_2009.sta); subroutine 'groupsFindings'

	unsorted index	Wariant	Gatunek	Pochodz	sorted means	Zmn6	Zmn7	Zmn8	Zmn9	Zmn10	1	2	3	4	5	6	7	8
1	2	kontrola	Alnus incana	Krzyż	178,38	1												
2	5	kontrola	Alnus incana	Rzepin	336,14	0,568532751												
3	9	kontrola	Alnus incana	Wolsztyn	532,31	0,201265868												
4	11	kontrola	Alnus glutinosa	Krzyż	550,35	0,179294972												
5	8	kontrola	Alnus glutinosa	Rzepin	626,76	0,105726068												
6	7	kontrola	Alnus glutinosa	Wolsztyn	650,2	0,0887744767												
7	12	KGHM	Alnus incana	Krzyż	712,89	0,0540065315												
8	10	KGHM	Alnus incana	Rzepin	771,93													
9	1	KGHM	Alnus incana	Wolsztyn	784,45													
10	3	KGHM	Alnus glutinosa	Krzyż	873,59													
11	6	KGHM	Alnus glutinosa	Rzepin	1001													
12	4	KGHM	Alnus glutinosa	Wolsztyn	1268,1													

Przypadek, gdy grupa jednorodna zawiera się całkowicie w poprzednio odnotowanej – jej kolumna zostanie usunięta

E:\GLP\Inne\ALNUS- K. Ufnalski\Alnus_2009\Alnus_BCF_BAF_TC_2009.sta Test NIR; zmienna Fe [ppm] (Alnus_BCF_BAF_TC_2009.sta); subroutine 'groupsFindings'

	unsorted index	Wariant	Gatunek	Pochodz	sorted means	Zmn6	Zmn7	Zmn8	Zmn9	Zmn10	1	2	3	4	5	6	7	8
1	2	kontrola	Alnus incana	Krzyż	178,38	1												
2	5	kontrola	Alnus incana	Rzepin	336,14	0,568532751	1											
3	9	kontrola	Alnus incana	Wolsztyn	532,31	0,201265868	0,478336111	1										
4	11	kontrola	Alnus glutinosa	Krzyż	550,35	0,179294972	0,438848211	0,947979304	1									
5	8	kontrola	Alnus glutinosa	Rzepin	626,76	0,105726068	0,293802384	0,73276746	0,782381558									
6	7	kontrola	Alnus glutinosa	Wolsztyn	650,2	0,0887744767	0,256657684	0,669984695	0,718137848									
7	12	KGHM	Alnus incana	Krzyż	712,89	0,0540065315	0,173781312	0,513959593	0,556867738									
8	10	KGHM	Alnus incana	Rzepin	771,93		0,115837201	0,386585815	0,423311232									
9	1	KGHM	Alnus incana	Wolsztyn	784,45		0,105768021	0,362275268	0,397600424									
10	3	KGHM	Alnus glutinosa	Krzyż	873,59		0,0527056517	0,217798691	0,243055454									
11	6	KGHM	Alnus glutinosa	Rzepin	1001			0,0908703458	0,103947506									
12	4	KGHM	Alnus glutinosa	Wolsztyn	1268,1													

Jeszcze 3 następne grupy jednorodne zostały usunięte, aż napotkano nową grupę, reprezentowaną ostatnią kolumną tej tabeli

E:\GLP\Inne\ALNUS- K. Ufnalski\Alnus_2009\Alnus_BCF_BAF_TC_2009.sta Test NIR; zmienna Fe [ppm] (Alnus_BCF_BAF_TC_2009.sta); subroutine 'groupsFindings'

	unsorted index	Wariant	Gatunek	Pochodz	sorted means	Zmn6	Zmn7	Zmn8	Zmn9	Zmn10	1	2	3	4	5	6	7	8
1	2	kontrola	Alnus incana	Krzyż	178,38	1												
2	5	kontrola	Alnus incana	Rzepin	336,14	0,568532751	1											
3	9	kontrola	Alnus incana	Wolsztyn	532,31	0,201265868	0,478336111	1										
4	11	kontrola	Alnus glutinosa	Krzyż	550,35	0,179294972	0,438848211	0,947979304										
5	8	kontrola	Alnus glutinosa	Rzepin	626,76	0,105726068	0,293802384	0,73276746										
6	7	kontrola	Alnus glutinosa	Wolsztyn	650,2	0,0887744767	0,256657684	0,669984695										
7	12	KGHM	Alnus incana	Krzyż	712,89	0,0540065315	0,173781312	0,513959593										
8	10	KGHM	Alnus incana	Rzepin	771,93		0,115837201	0,386585815	1									
9	1	KGHM	Alnus incana	Wolsztyn	784,45		0,105768021	0,362275268	0,96386939									
10	3	KGHM	Alnus glutinosa	Krzyż	873,59		0,0527056517	0,217798691	0,713231471									
11	6	KGHM	Alnus glutinosa	Rzepin	1001			0,0908703458	0,407722525									
12	4	KGHM	Alnus glutinosa	Wolsztyn	1268,1				0,0735816927									

Wszystkie następne grupy jednorodne zawierały się w poprzednio znalezionej

E:\GLP\Inne\ALNUS- K. Ufnalski\Alnus_2009\Alnus_BCF_BAF_TC_2009.sta Test NIR; zmienna Fe [ppm] (Alnus_BCF_BAF_TC_2009.sta); subroutine 'groupsFindings'

	unsorted index	Wariant	Gatunek	Pochodz	sorted means	Zmn6	Zmn7	Zmn8	Zmn9	Zmn10	1	2	3	4	5	6	7	8
1	2	kontrola	Alnus incana	Krzyż	178,38	1												
2	5	kontrola	Alnus incana	Rzepin	336,14	0,568532751	1											
3	9	kontrola	Alnus incana	Wolsztyn	532,31	0,201265868	0,478336111	1										
4	11	kontrola	Alnus glutinosa	Krzyż	550,35	0,179294972	0,438848211	0,947979304										
5	8	kontrola	Alnus glutinosa	Rzepin	626,76	0,105726068	0,293802384	0,73276746										
6	7	kontrola	Alnus glutinosa	Wolsztyn	650,2	0,0887744767	0,256657684	0,669984695										
7	12	KGHM	Alnus incana	Krzyż	712,89	0,0540065315	0,173781312	0,513959593										
8	10	KGHM	Alnus incana	Rzepin	771,93		0,115837201	0,386585815	1									
9	1	KGHM	Alnus incana	Wolsztyn	784,45		0,105768021	0,362275268	0,96386939									
10	3	KGHM	Alnus glutinosa	Krzyż	873,59		0,0527056517	0,217798691	0,713231471									
11	6	KGHM	Alnus glutinosa	Rzepin	1001			0,0908703458	0,407722525									
12	4	KGHM	Alnus glutinosa	Wolsztyn	1268,1				0,0735816927									

Wyniki grupowania po przycięciu tabeli do niezbędnego rozmiaru, symbolicznym opisanu grup i po przypisaniu średnim symboli grup

E:\GLP\Inne\ALNUS- K. Ufnalski\Alnus_2009\Alnus_BCF_BAF_TC_2009.sta Test NIR; zmienna Fe [ppm] (Alnus_BCF_BAF_TC_2009.sta); subroutine 'groupsFindings'

	unsorted index	Wariant	Gatunek	Pochodz	sorted means	A	B	C	D	Group
1	2	kontrola	Alnus incana	Krzyż	178,38	1				A
2	5	kontrola	Alnus incana	Rzepin	336,14	0,568532751	1			AB
3	9	kontrola	Alnus incana	Wolsztyn	532,31	0,201265868	0,478336111	1		ABC
4	11	kontrola	Alnus glutinosa	Krzyż	550,35	0,179294972	0,438848211	0,947979304		ABC
5	8	kontrola	Alnus glutinosa	Rzepin	626,76	0,105726068	0,293802384	0,73276746		ABC
6	7	kontrola	Alnus glutinosa	Wolsztyn	650,2	0,0887744767	0,256657684	0,669984695		ABC
7	12	KGHM	Alnus incana	Krzyż	712,89	0,0540065315	0,173781312	0,513959593		ABC
8	10	KGHM	Alnus incana	Rzepin	771,93		0,115837201	0,386585815	1	BCD
9	1	KGHM	Alnus incana	Wolsztyn	784,45		0,105768021	0,362275268	0,96386939	BCD
10	3	KGHM	Alnus glutinosa	Krzyż	873,59		0,0527056517	0,217798691	0,713231471	BCD
11	6	KGHM	Alnus glutinosa	Rzepin	1001			0,0908703458	0,407722525	CD
12	4	KGHM	Alnus glutinosa	Wolsztyn	1268,1				0,0735816927	D

Ostatnia tabela pokazuje, że decyzje o przynależności do grup zapadają przy różnych poziomach wiarygodności. To zainspirowało autora do podjęcia badań zależności wyników grupowania od poziomu istotności i do skonstruowania współczynnika oceniającego wyniki grupowania takiego współczynnika, który oceniałby niezależność postępowania od poziomu istotności. Współczynnik taki, jak to pokazały rozważania, ocenia zarazem zależność zmiennej zależnej od wektora zmiennych niezależnych, czyli jest współczynnikiem korelacji.

membership.

1. Najpierw pobiera wartość pustą z pierwszej komórki pierwszej wolnej kolumny arkusza „groups”.
2. Następnie wypełnia tymi wartościami całą przestrzeń działania.
3. W następnych krokach, w dwóch pętlach iteracyjnych sterowanych indeksami i oraz j , przeszukuje górny trójkąt MIR w porządku wskazywanym przez indeks sortujący w poszukiwaniu prawdopodobieństw większych lub równych zadanemu poziomowi istotności (α) włączając przekątną (przyjmuje się, że na przekątnej są wartości 1).
4. Jeśli taką znajdzie, to wpisuje tę wartość do k -tej kolumny na pozycji j . Indeks k wskazuje początkowo pierwszą wolną kolumnę i jest, w miarę potrzeby, stopniowo powiększany. Indeks j , to składnik pary (i, j) wskazującej położenie $(IdxVect(i), IdxVect(j))$ średniej, która z prawdopodobieństwem $1-\alpha$ nie różni się od aktualnego odniesienia, którym jest średnia na pozycji $(IdxVect(i), IdxVect(i+1))$, gdzie $IdxVect$ jest wektorem sortującym.
5. Oprócz tego sprawdza warunek *inc* zezwalający na powiększenie indeksu k przed kolejnym wykonaniem zewnętrznej pętli iteracyjnej sterowanej indeksem i .
6. Po przebiegnięciu i -tego wiersza powiększa indeks k jeśli warunek *inc* na to zezwala. W przeciwnym razie indeksu k nie powiększa i kasuje cyfrę 1 wpisaną początkowo na i -tej pozycji k -tej kolumny jako wartość z przekątnej MIR.
7. Zezwolenie na powiększenie indeksu k następuje, gdy w poprzedniej kolumnie, na j -tej pozycji komórka zawiera wartość $<\alpha$ (jest pusta) albo, gdy jest to pierwsza kolumna wykazu (bo kolumna ta jest poprzedzana niepustą kolumną średnich). Oznacza to, że aktualnie analizowana średnia należy do nowej grupy (por. rys. 1.).

Uwaga. W tabeli *groups* wypisujemy tylko wartości przekraczające próg α dla celów czysto wizualnych – aby upodobnić obraz zapisu do klasycznego zapisu „gwiazdkowego”.

Hierarchia czynników jest zgodna z kolejnością ich wybierania z listy.

SummaryOut

Zawartość arkusza wyników zależy od opcji.

Zaokrąglenia średnich

Stosujemy zasadę: błąd wyrażamy liczbą zawierającą zadaną ilość cyfr znaczących, czyli ze zbioru 1 do 9, a średnią podajemy z tą samą dokładnością, co błąd.

W publikacjach z dziedziny nauk przyrodniczych na ogół wymaga się wyrażania błędu liczbą z jedną cyfrą znaczącą. Ten sposób zaokrąglania poprawia czytelność wyników, a gdy błąd jest mniejszy od 1, daje także oszczędność miejsca w zapisie. Gdy błąd jest większy lub równy 10 nie ma oszczędności w rozmiarze zapisu. W tym przypadku możemy wybrać opcję zaokrąglania średniej i błędu do najbliższej liczby całkowitej.

Przykłady

Przy założeniu, że błąd wyrażamy **jedną** cyfrą znaczącą

wielkość	$0,3652 \pm 0,002619$	podajemy jako	$0,365 \pm 0,003$, a
wielkość	$3,652 \pm 0,02619$	podajemy jako	$3,65 \pm 0,03$, a
wielkość	$567,556 \pm 37,13$	podajemy jako	570 ± 40 , zaś

przyjmując **dwie** cyfry znaczące dla błędu, powyższe dane zapiszemy jako

$0,365 \pm 0,0026$ $3,65 \pm 0,026$ i 568 ± 37 odpowiednio.

Przy zaokrąglaniu do **najbliższej liczby całkowitej** (co dotyczyć może przypadków, gdy $\text{błąd} \geq 10$) otrzymamy 568 ± 37 , a pozostałe przykłady nie ulegną zmianie (zostanie zastosowane kryterium ilości cyfr znaczących dla błędu).

Zaokrąglenia prawdopodobieństw.

Prawdopodobieństwa zaokrąglamy do wskazanej ilości miejsc dziesiętnych po przecinku. Na ogół wymaga się, aby zaokrąślać je do trzech miejsc po przecinku. Niekiedy jednak oczekuje się, aby mniejsze prawdopodobieństwa podawać w formie wykładniczej, np. 0,0003217 jako $3,217 \text{E-}4$. Także i w tym przypadku należy zaokrąślać mantysę. My zaokrąglamy ją do liczby z jednym miejscem po przecinku, czyli przykładową liczbę sprowadzamy do postaci $3,2 \text{E-}4$. Właściwości arkuszy w Statistica 10 nie pozwalają jednak używać tej formy zapisu zgodnie z własnym życzeniem (wszystkie liczby większe od $1 \text{E-}6$ zapisuje w formie zwykłej) umożliwia więc stosowanie znaku wykrzyknika zamiast litery E. Wykrzyknik należy w dokumencie końcowym, zredagowanym w MS Word, zastąpić literą E.

Areas

ANO VA	spreadsheet	number of rows	number of columns
1	mainRes	$nUv(InList2)+BoBy(corrCoeff)+SATestNbr+2-free$	$InList2+(cols+InList2)*InList1$
	R_Trends	$alphaStepNbr+1$	$InList1+1$
	R_summaries	1	$InList2+InList1$
	Probabilities	$SATestNbr+2-free$	$InList2+InList1+1$
2	mainRes	$k1=nUv(InList2)+BoBy(corrCoeff)+SATestNbr+2$ $k1*nUv(InList2-1)+SATestNbr+4-free$	$InList2+(cols+InList2)*InList1$
	R_Trends	$alphaStepNbr+1$	$InList1*((nUv(InList2-1)+1))+1$
	R_summaries	$(nUv(InList2-1)+1)$	$InList2+InList1$
	Probabilities	$k1=SATestNbr+2$ $k1*nUv(InList2-1)+SATestNbr+4-free$	$InList2+InList1+1$
3	mainRes	$k2=nUv(InList2)+BoBy(corrCoeff)+SATestNbr+2$ $k2=k2*nUv(InList2-1)+SATestNbr+4$ $k2*nUv(InList2-2)+SATestNbr+8-free$	$InList2+(cols+InList2)*InList1$
	R_Trends	$alphaStepNbr+1$	$InList1*((nUv(InList2-1)+1)*nUv(InList2-2)+1)+1$
	R_summaries	$(nUv(InList2-1)+1)*nUv(InList2-2)+1$	$InList2+InList1$
	Probabilities	$k2=SATestNbr+2$ $k2=k2*nUv(InList2-1)+SATestNbr+4$ $k2*nUv(InList2-2)+SATestNbr+8-free$	$InList2+InList1+1$
4	mainRes	$k1=nUv(InList2)+BoBy(corrCoeff)+SATestNbr+2$ $k1=k1*nUv(InList2-1)+SATestNbr+4$ $k1=k1*nUv(InList2-2)+SATestNbr+8$ $k1*nUv(InList2-3)+SATestNbr+16-free$	$InList2+(cols+InList2)*InList1$
	R_Trends	$alphaStepNbr+1$	$InList1*(((nUv(InList2-1)+1)*nUv(InList2-2)+1)*nUv(InList2-3)+1)+1$
	R_summaries	$((nUv(InList2-1)+1)*nUv(InList2-2)+1)*nUv(InList2-3)+1$	$InList2+InList1$
	Probabilities	$k1=SATestNbr+2$ $k1=k1*nUv(InList2-1)+SATestNbr+4$ $k1=k1*nUv(InList2-2)+SATestNbr+8$ $k1*nUv(InList2-3)+SATestNbr+16-free$	$InList2+InList1+1$

$$free = (1 - BoBy(frts)) * \left[1 + \sum_{i=1}^{m-1} \prod_{j=i}^{m-1} nUv(m-j) \right] + (1 - BoBy(OWA)) * \prod_{i=1}^{m-1} nUv(m-i)^3, frts - \text{free term show}, OWA - \text{one way ANOVA } nUv - \text{number of unique values},$$

m , $InList2$ – number of category variables (factors), $InList1$ – number of dependent variables (features), $BoBy$ – boolean to byte conversion, $corrCoeff$ – calculate correlation coefficient, $SATestNbr$ – number of simple ANOVA tests which will be done (i.e. some or all of F-Snedecore, Welch, Levene, Brown-Forsythe), $cols$ – number of columns for results – features means, standard errors and repetitions counts; $cols=2$ or 4 ; 2 – when means and errors are concatenated, 4 – otherwise, $alphaStepNbr$ – number of significance levels.

Includes

ANOVA	formula	loop control variable	
1	includes=""	–	
2	includes="(" + namevarlist2(InList2-1) + "" = " + Str(valuesArr(InList2-1,k1)) + ")"	↑	k1
	includes=""	↓	
3	t_includes="(" + namevarlist2(InList2-2) + "" = " + Str(valuesArr(InList2-2,k2)) + ")"		k2
	includes=t_includes + " and (" + namevarlist2(InList2-1) + "" = " + Str(valuesArr(InList2-1,k1)) + ")"	↑	k1
	includes=t_includes	↓	k2
	includes=""	↓	–
4	t2_includes="(" + namevarlist2(InList2-3) + "" = " + Str(valuesArr(InList2-3,k3)) + ")"	↑	k3
	t1_includes=t2_includes + " and (" + namevarlist2(InList2-2) + "" = " + Str(valuesArr(InList2-2,k2)) + ")"	↑	k2
	includes=t1_includes + " and (" + namevarlist2(InList2-1) + "" = " + Str(valuesArr(InList2-1,k1)) + ")"	↑	k1
	includes=t1_includes	↓	k2
	includes=t2_includes	↓	k3
	includes=""	↓	–

namevarlist2 – names of category variables (factors), valuesArr – powers of sets of category variables (manyess), ↑ – loop beginning, ↓ – loop end.

³ W implementacji pierwszego członu skorzystano z postaci rekurencyjnej $free(m) = (1 - BoBy(frts)) * \{ [free(i-1) * nUv(m-i)] + 1 \}$, $free(0) = 1$, $i = 1 \dots m-1$

O programie

Program otwieramy w środowisku STATISTICA 10. Uruchomienie następuje po naciśnięciu klawisza F5.

Uruchomienie programu musi poprzedzać wczytanie arkusza danych.

Po uruchomieniu program m.in. wczytuje dane z pliku *MultiANOVAinit.txt*. Dane te zawierają wykaz ilości i numerów ostatnio wybranych zmiennych na liście zmiennych zależnych (cech) i na liście zmiennych klasyfikujących (czynników). Zmienne o tych samych numerach będą wybrane także i w bieżącej sesji (jako propozycja). Plik *MultiANOVAinit.txt* jest tworzony automatycznie w każdej sesji i kierowany jest do tego samego katalogu, w którym jest plik programu *MultiANOVA.svb*. Zmienne z listy zmiennych zależnych są opcjonalnie sortowane wg kolejności numerów na liście. Ma to ułatwić zachowanie oryginalnej kolejności na wydrukach wyników analizy po dodaniu nowych zmiennych z tej listy w powtórzeniach analiz (przy