

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
Кафедра штучного інтелекту

Звіт з виконання завдань
комп'ютерного практикуму № 7
Кореляційно-регресійний аналіз у середовищі STATISTICA
з кредитного модуля «Багатовимірний статистичний аналіз»

Звіт склав
студент гр. КІ-01
Копцов В.О.
Прийняла: *Ірина Джигирей*

Мета роботи. Опанувати підходи та процедури розрахунку індексів, набути вмінь та досвіду проведення агрегування показників з використанням відновлення даних.

Завдання. За наданими значеннями показників згідно індивідуального завдання (Додаток А) виконати аналіз індикаторів та процедуру агрегування. Розрахунки провести у застосунку Statistica.

Варіант 30

Значення показників розміщено у файлі: msa-cr07-2019_k-z.sta

Метод відновлення даних: безумовне відновлення (медіана)

Кількість кластерів: 5

Нормалізування: від 0 до 1

Агрегування суб-індексів: сума

Агрегування індексу: незважене арифметичне середнє

Хід виконання завдань практикуму

Завантажив дані країн з k до z

Data: msa-cp07-2019_k-z (10v by 79c)										
	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
KAZ	8820	5,04	99,99	4,80	73,18	81,88	29,721	10,882	1,70	11,457
KEN	1830	0,27	35,31	5,01	66,70	22,57	4,593	10,191	68,08	0,424
KGZ	1240	4,88	104,41	6,92	71,60	51,00	6,738	6,712	27,88	1,557
KHM	1530	17,43	60,06	0,15	69,82		1,194	21,958	53,36	0,981
KOR	33830	17,68		3,75	83,23	96,16	32,404	13,975	3,36	11,799
KWT	36200	12,16		2,17	75,49	99,54	0,903	0,449	0,06	22,022
LAO	2540		93,96	0,85	67,92	30,84	20,609	5,614	48,43	2,610
LBN	7280	-23,83	143,94	11,35	78,93	80,90	3,891	13,226	5,46	4,077
LBR	610		43,71	3,02	64,10	22,00		5,191	87,24	0,239
LCA	10840		31,88	15,32	76,20	50,70	10,527	4,377	9,75	2,954
LKA	4010	18,14	68,79	4,35	76,98	29,00	1,102	22,171	49,91	1,091
LSO	1330	2,71	34,79	22,44	54,33	42,30	0,253	4,545	39,47	0,362
LTU	19060	11,58		6,26	76,28	81,58	12,030	35,323	33,53	4,198
LUX	77110	9,69		5,59	82,64	97,12	6,573	24,067	16,45	15,306
LVA	17810	4,04		6,31	75,39	86,14	17,217	21,197	41,52	3,955
MAR	3200	19,80	46,85	9,28	76,68	74,38	4,852	15,458	10,69	1,960
MDA	4580	4,34	59,74	5,10	71,90		2,961	51,387	22,04	3,325
MDG	520	1,72	29,90	1,86	67,04		0,366	5,156	82,77	0,153
MDV	9640	10,76	53,07	5,34	78,92	62,50		13,000	1,11	3,974
MEX	9470	7,32	38,33	3,48	75,05	70,07	20,406	9,980	10,34	3,522
MKD	5890	18,25	74,73	17,26	76,60	81,41	4,217	16,614	16,32	3,997
MLI	860		30,62	7,44	59,31	24,20	5,480	5,254	76,64	0,297
MMR	1270	20,28	16,75	0,50	67,13	34,90	2,763	16,838	57,85	0,679
MNE	9140		148,79	15,12	76,68	73,48	6,890	0,669	38,74	4,180
MNG	3850	-0,21	250,51	5,44	69,87	51,08	18,944	0,853	3,32	7,153

Відновлюємо дані за допомогою безумовного медіанного відновлення.

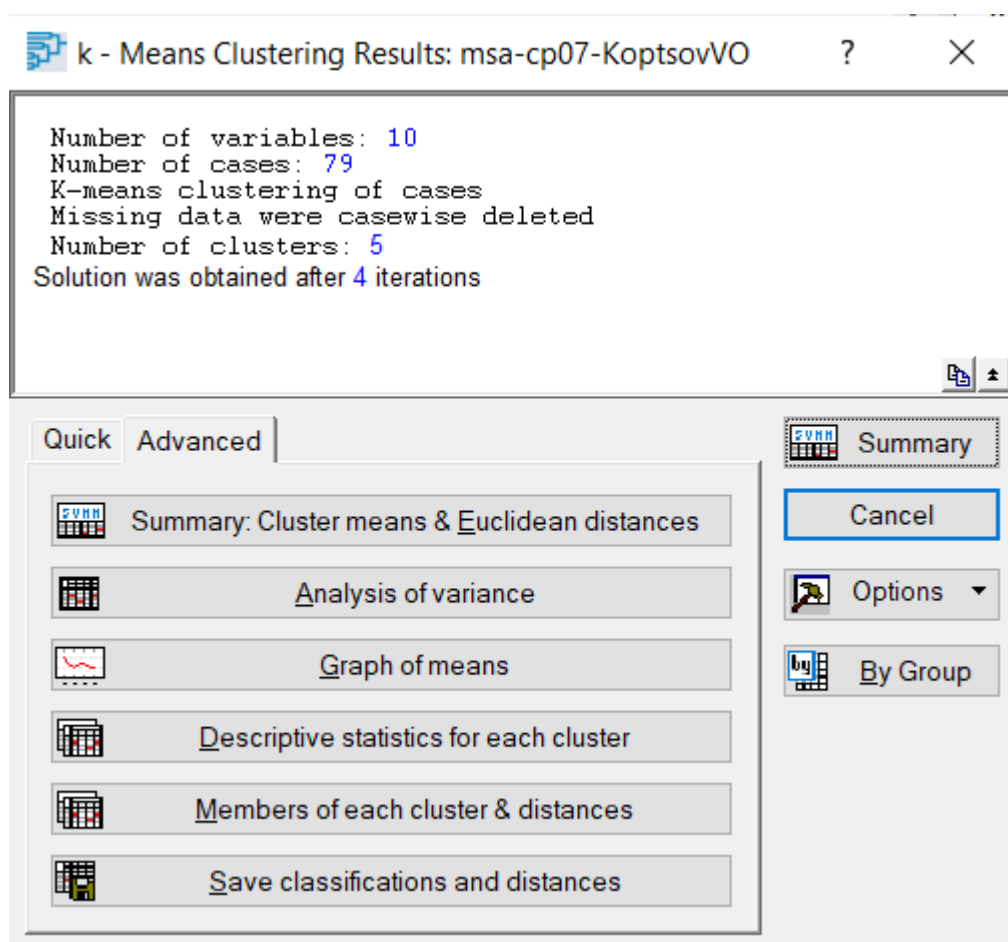
Таблиця кореляцій

	Correlations (msa-cp07-KoptsovVO)											
	Marked correlations are significant at p < ,05000											
	N=79 (Casewise deletion of missing data)											
Variable	Means	Std.Dev.	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
x1	12570,13	18756,96	1,000	0,217	-0,081	-0,170	0,635	0,661	0,277	-0,150	-0,334	0,669
x2	9,68	9,84	0,217	1,000	-0,124	-0,111	0,210	0,083	0,077	-0,144	0,052	0,072
x3	58,81	38,29	-0,081	-0,124	1,000	0,125	0,129	0,011	-0,034	-0,139	-0,188	0,025
x4	6,53	5,66	-0,170	-0,111	0,125	1,000	-0,187	-0,003	-0,234	-0,162	-0,094	-0,156
x5	71,88	7,51	0,635	0,210	0,129	-0,187	1,000	0,781	0,340	-0,051	-0,653	0,513
x6	58,70	26,81	0,661	0,083	0,011	-0,003	0,781	1,000	0,272	-0,124	-0,677	0,677
x7	9,59	12,13	0,277	0,077	-0,034	-0,234	0,340	0,272	1,000	-0,078	-0,295	0,152
x8	16,01	13,86	-0,150	-0,144	-0,139	-0,162	-0,051	-0,124	-0,078	1,000	0,082	-0,196
x9	36,14	28,73	-0,334	0,052	-0,188	-0,094	-0,653	-0,677	-0,295	0,082	1,000	-0,602
x10	4,22	5,50	0,669	0,072	0,025	-0,156	0,513	0,677	0,152	-0,196	-0,602	1,000

x1, x5, x6, x10 між собою мають помітну залежність. Всі інші між собою ніяк не пов'язані.

Це валовий національний дохід на душу населення, очікувана тривалість життя при народженні, використання інтернету, емісія CO₂. Ці показники логічно також корелюють, але в інших змінних не спостерігається кореляції, хоча вони також можуть мати деяку кореляцію. На мою думку це може бути спричинено паганим методом відновлення даних. Хоча ще є такий випадок як Філіппіни в яких дуже високий відсоток високотехнологічного експорту стикається з доволі низьким валовим національним доходом на душу населення.

Кластерний аналіз(кількість кластерів - 5)



Описові статистики і члени (кластеру) для кожного кластеру

1 кластер

Variable	Cluster contains 7 cases		
	Mean	Standard Deviation	Variance
x1	64805,71	10802,18	116687200
x2	17,58	7,90	62
x3	48,52	0,00	0
x4	3,77	2,11	4
x5	81,92	1,76	3
x6	94,42	4,16	17
x7	19,58	16,52	273
x8	11,77	11,95	143
x9	21,65	25,36	643
x10	12,76	9,67	94

	Distance
LUX	3890,962
NLD	3660,570
NOR	5332,973
QAT	1146,614
SGP	2133,223
SWE	2651,817
USA	368,223

2 кластер

Variable	Mean	Standard Deviation	Variance
x1	37693,33	4787,968	22924630
x2	14,29	2,971	9
x3	48,52	0,000	0
x4	3,34	1,034	1
x5	80,14	4,100	17
x6	95,30	4,730	22
x7	14,39	16,231	263
x8	5,50	7,385	55
x9	11,02	16,215	263
x10	13,55	7,746	60

	Cluster contains 3 cases
	Distance
KOR	1221,713
KWT	472,281
NZL	1693,940

3 кластер

Variable	Cluster contains 10 cases		
	Mean	Standard Deviation	Variance
x1	19300,00	3693,309	13640530
x2	8,62	9,391	88
x3	60,86	39,022	1523
x4	5,36	1,928	4
x5	77,99	2,220	5
x6	82,25	8,576	74
x7	8,32	4,417	20
x8	16,18	13,204	174
x9	23,36	18,646	348
x10	6,81	4,770	23

	Distance
LTU	76,325
LVA	471,247
OMN	907,656
PAN	1385,547
POL	1245,965
PRT	1233,300
SAU	1239,667
SVK	16,836
SVN	2112,407
URY	483,996

4 кластер

Variable	Mean	Standard Deviation	Variance
x1	9160,625	2189,619	4794433
x2	5,976	9,528	91
x3	66,897	39,401	1552
x4	9,206	7,308	53
x5	75,394	3,448	12
x6	71,834	9,683	94
x7	12,476	13,158	173
x8	15,945	12,753	163
x9	13,841	10,618	113
x10	5,341	2,941	9

	Distance
KAZ	108,515
LBN	595,301
LCA	531,242
MDV	151,773
MEX	98,327
MKD	1034,284
MNE	28,406
MUS	1163,650
MYS	664,059
PER	733,985
ROU	1093,997
RUS	670,337
SRB	670,625
THA	604,339
TUR	167,497
ZAF	787,668

5 кластер

Variable	Mean	Standard Deviation	Variance
x1	2017,442	1387,116	1924091
x2	9,704	10,090	102
x3	57,718	42,312	1790
x4	6,471	5,813	34
x5	66,950	6,175	38
x6	39,964	20,424	417
x7	6,852	11,241	126
x8	17,425	15,012	225
x9	51,520	27,624	763
x10	1,164	1,387	2

	Distance
KEN	60,301
KGZ	246,462
KHM	154,428
LAO	165,782
LBR	445,293
LKA	630,140
LSO	217,698
MAR	374,383
MDA	810,526
MDG	473,822
MLI	366,264
MMR	236,765
MNG	582,962
MOZ	483,780
MRT	94,863
MWI	461,213
NAM	996,978
NER	451,695
NGA	19,125
NIC	42,306
NPL	249,603
PAK	126,272
PHL	579,957
PRY	1104,498

RWA	375,772
SDN	340,971
SEN	182,696
SLE	470,588
SLV	620,746
STP	22,742
SWZ	535,526
TCD	416,942
TGO	347,474
TJK	299,809
TLS	24,046
TUN	507,175
TZA	290,490
UGA	391,667
UKR	409,357
UZB	50,744
VNM	399,662
ZMB	187,369
ZWE	255,618

Загалом всередині кожного кластеру спостерігається якась логіка. Цікаво тільки як нова зеландія попала до кувейту і північної Кореї.

Дані по x3 явно були відновленні для країн перших двох кластерів(там нема дисперсії з доволі високим середнім значенням)

Описові статистики.

Variable	Descriptive Statistics (msa-cp07-KoptsovVO)				
	Valid N	Mean	Minimum	Maximum	Std.Dev.
x1	79	12570,13	490,0000	81670,00	18756,96
x2	79	9,68	-23,8272	35,28	9,84
x3	79	58,81	7,4479	250,51	38,29
x4	79	6,53	0,1000	28,47	5,66
x5	79	71,88	54,2390	83,60	7,51
x6	79	58,70	9,8000	99,65	26,81
x7	79	9,59	0,0025	62,25	12,13
x8	79	16,01	0,2515	56,82	13,86
x9	79	36,14	0,0000	90,22	28,73
x10	79	4,22	0,0778	32,47	5,50

Нормалізував дані

Повторний кластерний аналіз

Кластер 1

	Distance
KGZ	0,052719
KHM	0,165820
LAO	0,149277
LCA	0,141853
LKA	0,155363
MNE	0,161306
MNG	0,267250
MRT	0,138242
NAM	0,169784
NIC	0,108611
PAN	0,179816
PER	0,118626
PRY	0,139316
SEN	0,105309
STP	0,112354
TJK	0,078985
TLS	0,186569
ZAF	0,272029

Кластер 2

	Distance
KOR	0,127970
KWT	0,169073
LUX	0,143421
NLD	0,134548
NOR	0,211381
NZL	0,105159
QAT	0,229091
SAU	0,174521
SGP	0,212893
SWE	0,166154
USA	0,095424

Кластер 3

	Distance
KAZ	0,149406
LBN	0,224150
LTU	0,108822
LVA	0,105403
MAR	0,107520
MDA	0,184677
MDV	0,102844
MEX	0,092433
MKD	0,151645
MUS	0,138022
MYS	0,236742
OMN	0,214759
PHL	0,297418
POL	0,104147
PRT	0,115062
ROU	0,104964
RUS	0,130844
SLV	0,132799
SRB	0,081834
SVK	0,065333
SVN	0,127392
THA	0,123196
TUN	0,128262
TUR	0,105037
UKR	0,215745
URY	0,178666
UZB	0,135650
VNM	0,151007

Кластер 4

	Distance
KEN	0,090198
LBR	0,083061
LSO	0,237370
MDG	0,158436
MLI	0,061801
MMR	0,128446
MOZ	0,154630
NER	0,153798
NPL	0,167168
SDN	0,151157
SWZ	0,222277
TCD	0,147478
TZA	0,117850
ZMB	0,141648
ZWE	0,100497

Кластер 5

	Distance
MWI	0,063760
NGA	0,114547
PAK	0,122509
RWA	0,115035
SLE	0,154952
TGO	0,075252
UGA	0,076858

Результати кластеризування значно змінилися. З позитивних моментів, які необхідно помітити тепер усі розвинені(багаті) країни вийшли в один кластер(другий), а Україна нарешті пішла від своїх Африканських друзів в компанію Польщі і Туреччини єдине припущення, що Україні допоміг найбільший відсоток орних земель.

Створив суб-індекси
Кластеризував за ними

Members of Cluster Number 1 (msa-cp07-KoptsovVO) and Distances from Respective Cluster Center Cluster contains 18 cases	
	Distance
KGZ	0,100085
KHM	0,241058
LAO	0,147959
LKA	0,236341
LSO	0,208261
MNG	0,388079
MRT	0,264618
MUS	0,241270
NAM	0,152257
NIC	0,163497
PER	0,193160
SDN	0,219015
SEN	0,231079
SLV	0,169150
STP	0,182389
TJK	0,144815
TLS	0,450728
UZB	0,226541

Members of Cluster Number 2 (msa-cp07-KoptsovVO) and Distances from Respective Cluster Center Cluster contains 12 cases	
	Distance
LTU	0,13955
LVA	0,14594
MDA	0,25559
POL	0,06860
PRY	0,10988
ROU	0,05024
SRB	0,08098
SVK	0,13139
SWZ	0,18913
THA	0,03320
UKR	0,16293
URY	0,15066

Members of Cluster Number 3 (msa-cp07-KoptsovVO) and Distances from Respective Cluster Center Cluster contains 10 cases	
	Distance
KOR	0,26623
LUX	0,19545
NLD	0,09729
NOR	0,20001
NZL	0,20478
PAN	0,38964
QAT	0,28989
SGP	0,43104
SWE	0,07734
USA	0,11361

Members of Cluster Number 4 (msa-cp07-KoptsovVO) and Distances from Respective Cluster Center Cluster contains 20 cases	
	Distance
KAZ	0,073558
KWT	0,236947
LBN	0,163717
LCA	0,164433
MAR	0,063269
MDV	0,196972
MEX	0,143976
MKD	0,169092
MNE	0,220561
MYS	0,337139
OMN	0,200579
PHL	0,091008
PRT	0,066543
RUS	0,139203
SAU	0,181194
SVN	0,117599
TUN	0,055821
TUR	0,134474
VNM	0,140223
ZAF	0,158079

Members of Cluster Number 5 (msa-cp07-KoptsovVO) and Distances from Respective Cluster Center Cluster contains 19 cases	
	Distance
KEN	0,197604
LBR	0,085970
MDG	0,307158
MLI	0,153438
MMR	0,165170
MOZ	0,221908
MWI	0,164402
NER	0,208308
NGA	0,246137
NPL	0,256877
PAK	0,058960
RWA	0,310010
SLE	0,303734
TCD	0,358270
TGO	0,315842
TZA	0,148861
UGA	0,257438
ZMB	0,343550
ZWE	0,094781

Знову нові результати. Тепер “розвинутий” кластер - 3. Україна потрапила в один кластер з Польщею, Латвією і Литвою, що, як на мене, гарний результат.

Сформував індекс

Перші 20 країн за значенням індексу

	21 C1	22 C2	23 C3	24 I
NOR	1,824	2,448	0,931	1,734
LUX	1,680	2,238	1,073	1,664
SGP	1,760	2,819	0,273	1,617
SWE	1,595	2,396	0,795	1,595
NLD	1,536	2,364	0,879	1,593
QAT	1,749	1,885	1,028	1,554
KOR	1,282	2,598	0,642	1,507
USA	1,488	2,148	0,866	1,501
LTU	0,997	1,960	1,119	1,359
NZL	1,317	2,130	0,569	1,339
MNE	1,253	2,113	0,563	1,310
MKD	1,055	2,231	0,591	1,293
URY	0,870	2,068	0,935	1,291
LVA	0,854	2,065	0,950	1,290
POL	0,942	1,863	1,007	1,270
PAN	1,652	1,677	0,434	1,254
SVK	0,915	1,969	0,861	1,248
TUR	0,937	2,040	0,746	1,241
KWT	1,218	1,810	0,682	1,236
SVN	1,084	2,017	0,584	1,228

Останні 20 країн за значенням індексу

	21 C1	22 C2	23 C3	24 I
MWI	0,628	0,658	1,480	0,922
KGZ	0,894	1,398	0,469	0,920
NGA	0,542	0,601	1,592	0,912
LAO	0,946	1,058	0,710	0,904
MDG	0,525	1,137	1,006	0,889
PER	0,748	1,505	0,397	0,884
UZB	0,788	1,484	0,280	0,851
MMR	0,794	0,777	0,953	0,841
NAM	0,742	1,369	0,403	0,838
PAK	0,605	0,683	1,201	0,830
MRT	1,132	1,026	0,301	0,820
SEN	0,821	0,909	0,719	0,817
LBR	0,715	0,663	1,059	0,812
ZWE	0,580	0,664	1,104	0,783
MOZ	0,892	0,419	0,993	0,768
KEN	0,539	0,813	0,941	0,764
MLI	0,665	0,680	0,945	0,763
LSO	0,572	1,156	0,522	0,750
SLE	0,303	0,342	1,221	0,622
TCD	0,659	0,124	0,933	0,572

Висновки:

В процесі виконання комп'ютерного практикуму №7 я виконав розрахунки згідно індивідуального завдання і опанував підходи та процедури розрахунку індексів, набув вмінь та досвіду проведення агрегування показників з використанням відновлення даних. Виконання цього комп'ютерного практикуму було дуже зручним завдяки вбудованим в Statistica інструментам. Я багато разів виконував кластерний аналіз і майже миттєво отримував результати. Можливо єдине, що мені було потрібно, це використання декількох таблиць, але, наскільки я розумію, такий функціонал є в Statistica.