

Mat-2.3112 Tilastolliset monimuuttujamenetelmät

Tentti 15.5.2008 / Mellin

Kirjoita *selvästi jokaiseen koepaperiin* alla mainitussa järjestyksessä:

- Mat-2.3112 Monim / Tentti 15.5.2008
- opiskelijanumero + kirjain
- TEKSTATEN sukunimi, etunimet
- koulutusohjelma, vuosikurssi
- mahdolliset entiset nimet ja koulutusohjelmat
- nimikirjoitus

Sallitut apuvälineet: laskin (Mellinin taulukko- ja kaavakokoelma saa olla mukana – tosin siitä ei liene hyötyä).

Vastaa lyhyesti ja ytimekkäästi, mutta esitä vastauksissasi niin paljon perusteluita, että käy selvästi ilmi mitä tarkoitat.

Kaavojen käyttö on suotavaa.

1. Multinormaalijakauma

Esittele multinormaalijakauman keskeisimmät ominaisuudet. Käsittele ainakin seuraavia asioita:

- (a) Multinormaalijakauman tiheysfunktio ja sen geometria.
- (b) Multinormaalijakauman karakterisointi.
- (c) Multinormaalijakauman reunajakaumat.
- (d) Multinormaalijakauman ehdolliset jakaumat.
- (e) Riippumattomuus, korreloimattomuus ja multinormaalijakauma.
- (f) Multinormaalijakaumaa noudattavien satunnaismuuttujien neliömuotojen jakaumat.

Huomaa, että parametrien estimointia ja testausta ei tarvitse käsitellä.

2. Faktorianalyysi

Seuraavilla sivuilla on esitetty Ness-ohjelman tulostus eräästä faktorianalyysistä.

Aineistona on 48 parasta kymmenottelijaa vuodelta 1973. Muuttujina ovat ottelijoiden tulokset (pisteiksi muunnettuna) kymmenottelun 10:stä lajimuuttujasta (100 m, Pituushyppy, Kuula, Korkeushyppy, 400 m, 110 m:n aidat, Kiekko, Seiväs, Keihäs, 1500 m).

- (a) Esittele faktorianalyysin perusidea ja tavoite sekä faktorianalyysin malli ja sen osat.
- (b) Kuinka monta faktoria valitsisit tulostuksen perusteella? Miksi?
- (c) Esitä valitsemillesi faktoreille tulkinnat.
- (d) Mikä on muuttujan kuula ja 2. faktorin korrelaatio?
- (e) Selitä tulostuksen taulukot

Factor Loadings after Varimax Rotation

ja

Communalities after Varimax Rotation

- (f) Faktoreille on tulostuksen mukaan siis tehty varimax-rotatio, Mitä tällä tarkoitetaan ja miksi se on tehty?

Factor Analysis Report

Database

KYMMENOT.S0

Correlation Section

Variables	J100m	Pituush	Kuula	Korkeush	J400m
J100m	1.000000	0.171985	-0.027952	-0.411700	0.456082
Pituush	0.171985	1.000000	-0.034393	-0.003325	0.133463
Kuula	-0.027952	-0.034393	1.000000	0.162542	-0.303707
Korkeush	-0.411700	-0.003325	0.162542	1.000000	-0.338827
J400m	0.456082	0.133463	-0.303707	-0.338827	1.000000
Jaidat	0.315991	0.298065	0.086498	-0.038964	0.175548
Kiekkko	0.014343	0.020875	0.727330	0.216995	-0.344647
Seivas	0.054723	0.061024	-0.204229	-0.117765	0.006584
Keihas	-0.221250	0.153750	0.023140	0.149779	-0.104689
J1500m	-0.291744	-0.206665	-0.446248	-0.146150	0.302178

Variables	Jaidat	Kiekkko	Seivas	Keihas	J1500m
J100m	0.315991	0.014343	0.054723	-0.221250	-0.291744
Pituush	0.298065	0.020875	0.061024	0.153750	-0.206665
Kuula	0.086498	0.727330	-0.204229	0.023140	-0.446248
Korkeush	-0.038964	0.216995	-0.117765	0.149779	-0.146150
J400m	0.175548	-0.344647	0.006584	-0.104689	0.302178
Jaidat	1.000000	0.047707	-0.073497	-0.148188	-0.224552
Kiekkko	0.047707	1.000000	-0.181800	0.135561	-0.573501
Seivas	-0.073497	-0.181800	1.000000	-0.128523	0.012496
Keihas	-0.148188	0.135561	-0.128523	1.000000	-0.065414
J1500m	-0.224552	-0.573501	0.012496	-0.065414	1.000000

Phi=0.243776 Log(Det|R|)=-2.711129 Bartlett Test=116.13 DF=45 Prob=0.000000

Eigenvalues after Varimax Rotation

No.	Eigenvalue	Individual Percent	Cumulative Percent	Scree Plot
1	2.036656	41.75	41.75	
2	1.371836	28.12	69.88	
3	0.970867	19.90	89.78	
4	0.515542	10.57	100.35	
5	0.200662	4.11	104.46	
6	0.116739	2.39	106.86	
7	0.002351	0.05	106.90	
8	-0.035777	-0.73	106.17	
9	-0.087954	-1.80	104.37	
10	-0.213005	-4.37	100.00	

Eigenvectors after Varimax Rotation

Variables	Factor1	Factor2	Factor3	Factor4
J100m	-0.110444	-0.682761	0.217460	-0.159580
Pituush	0.019363	-0.267930	-0.716736	0.192549
Kuula	0.497530	-0.093440	0.325838	0.278261
Korkeush	0.237250	0.259567	-0.273780	0.077897
J400m	-0.373744	-0.341373	0.081483	0.518043
Jaidat	0.029197	-0.344775	-0.213009	0.142418
Kiekkko	0.560863	-0.119344	0.188568	0.206521
Seivas	-0.123064	-0.023066	-0.085754	-0.505672
Keihas	0.109241	0.136339	-0.332768	0.273435
J1500m	-0.448799	0.342310	0.228507	0.438394

Factor Loadings after Varimax Rotation

Variables	Factor1	Factor2	Factor3	Factor4
J100m	0.150974	-0.803099	-0.376639	-0.081963
Pituush	-0.011014	0.081202	-0.657499	-0.027176
Kuula	0.711639	0.061311	0.063426	0.396451
Korkeush	0.179120	0.496907	0.021564	0.062895
J400m	-0.462695	-0.532637	-0.276283	0.216272
Jaidat	0.109352	-0.206409	-0.414084	0.052732
Kiekkö	0.799911	0.127580	-0.029744	0.343864
Seivas	-0.087281	-0.080227	0.005980	-0.364742
Keihas	0.007616	0.347238	-0.131628	0.129591
J1500m	-0.775660	-0.002264	0.355210	0.171259

Communalities after Varimax Rotation

Variables	Factor1	Factor2	Factor3	Factor4	Communality
J100m	0.022793	0.644968	0.141857	0.006718	0.816336
Pituush	0.000121	0.006594	0.432305	0.000739	0.439758
Kuula	0.506430	0.003759	0.004023	0.157173	0.671385
Korkeush	0.032084	0.246916	0.000465	0.003956	0.283421
J400m	0.214087	0.283702	0.076332	0.046774	0.620895
Jaidat	0.011958	0.042605	0.171465	0.002781	0.228808
Kiekkö	0.639858	0.016277	0.000885	0.118242	0.775262
Seivas	0.007618	0.006436	0.000036	0.133037	0.147127
Keihas	0.000058	0.120574	0.017326	0.016794	0.154752
J1500m	0.601649	0.000005	0.126174	0.029330	0.757158

Factor Structure Summary after Varimax Rotation

Factor1	Factor2	Factor3	Factor4
Kiekkö	J100m	Pituush	
J1500m	J400m	Jaidat	
Kuula	Korkeush		
J400m			

Score Coefficients after Varimax Rotation

Variables	Factor1	Factor2	Factor3	Factor4
J100m	0.2000296	-0.6024989	-6.204705E-02	-0.166915
Pituush	-0.1739525	0.3923906	-0.8972586	2.863017E-02
Kuula	0.2864008	-0.1324258	0.1755287	0.5878128
Korkeush	-1.038744E-02	0.4170868	-0.1727713	4.800342E-02
J400m	-0.3354649	-0.2807662	-0.2501171	0.7276023
Jaidat	-0.012424	-2.084213E-02	-0.4232512	0.1360313
Kiekkö	0.3264409	-4.219654E-02	5.088198E-02	0.4443399
Seivas	0.1116862	-4.763842E-02	7.328439E-02	-0.7792428
Keihas	-0.1515322	0.3824744	-0.352806	0.2926772
J1500m	-0.4665062	-1.787738E-02	0.2521003	0.6514261

3. Erotteluanalyysi

Seuraavilla sivuilla on esitetty Ncss-ohjelman tulostus eräästä erotteluanalyysistä.

Analysoitavana on kolmea iiris-kasvin muotoa (Iris Setosa, Iris Versicolor, Iris Virginica) koskeva aineisto. Muuttujina on ko. iiris-lajien kukkien teriön erään terälehdén pituus ja leveys (PetalLenght, PetalWidth) ja verhiön erään verholehdén pituus ja leveys (SepalLenght, SepalWidth).

- (a) Esittele erotteluanalyysin perusidea ja tavoite sekä erotteluanalyysin malli ja sen osat.
- (b) Esitä erotteluanalyysille geometrinen tulkinta.
- (c) Kuinka monta erottelumuuttujaa valitset tulostuksen perusteella? Miksi?
- (d) Ovatko kaikki muuttujat merkityksellisiä erotteluinformaation kannalta?
- (e) Kuinka hyvin erottelumuuttujat ovat onnistuneet erottamaan lajit toisistaan?
- (f) Mitkä ovat erotteluanalyysin ja *kanonisen analyysin* yhteydet?

Discriminant Analysis Report

Database MMLU\HARJ\FISHER.S0
 Dependent Iris

Group Means

Variable	Iris			Overall
	Setosa	Versicolor	Virginica	
SepalLength	50.06	59.36	65.88	58.43333
SepalWidth	34.28	27.7	29.74	30.57333
PetalLength	14.62	42.6	55.52	37.58
PetalWidth	2.46	13.26	20.26	11.99333
Count	50	50	50	150

Group Standard Deviations

Variable	Iris			Overall
	Setosa	Versicolor	Virginica	
SepalLength	3.524897	5.161712	6.358796	8.280662
SepalWidth	3.790644	3.137983	3.224966	4.358663
PetalLength	1.73664	4.69911	5.518947	17.65298
PetalWidth	1.053856	1.977527	2.7465	7.622377
Count	50	50	50	150

Total Correlation\Covariance

Variable	Variable			
	SepalLength	SepalWidth	PetalLength	PetalWidth
SepalLength	68.56935	-4.243401	127.4315	51.62707
SepalWidth	-0.117570	18.99794	-32.96564	-12.16394
PetalLength	0.871754	-0.428440	311.6278	129.5609
PetalWidth	0.817941	-0.366126	0.962865	58.10063

Between-Group Correlation\Covariance

Variable	Variable			
	SepalLength	SepalWidth	PetalLength	PetalWidth
SepalLength	3160.607	-997.6334	8262.42	3563.967
SepalWidth	-0.745075	567.2466	-2861.98	-1146.633
PetalLength	0.994135	-0.812838	21855.14	9338.7
PetalWidth	0.999768	-0.759258	0.996232	4020.667

Within-Group Correlation\Covariance

Variable	Variable			
	SepalLength	SepalWidth	PetalLength	PetalWidth
SepalLength	26.50082	9.272109	16.75143	3.840136
SepalWidth	0.530236	11.53878	5.524354	3.27102
PetalLength	0.756164	0.377916	18.51878	4.266531
PetalWidth	0.364506	0.470535	0.484459	4.188163

Variable Influence Section

Variable	Removed Lambda	Removed F-Value	Removed F-Prob	Alone Lambda	Alone F-Value	Alone F-Prob	R-Squared Other X's
SepalLength	0.938463	4.72	0.010329	0.381294	119.26	0.000000	0.858612
SepalWidth	0.766480	21.94	0.000000	0.599217	49.16	0.000000	0.524007
PetalLength	0.669206	35.59	0.000000	0.058628	1180.16	0.000000	0.968012
PetalWidth	0.743001	24.90	0.000000	0.071117	960.01	0.000000	0.937850

Linear Discriminant Functions

Variable	Iris		
	Setosa	Versicolor	Virginica
Constant	-85.20985	-71.754	-103.2697
SepalLength	2.354417	1.569821	1.244585
SepalWidth	2.358787	0.707251	0.3685279
PetalLength	-1.643064	0.5211451	1.276654
PetalWidth	-1.739841	0.6434229	2.107911

Regression Coefficients

Variable	Iris		
	Setosa	Versicolor	Virginica
Constant	0.1182229	1.577059	-0.6952819
SepalLength	6.602977E-03	-2.015369E-03	-4.587608E-03
SepalWidth	2.428479E-02	-4.456162E-02	2.027684E-02
PetalLength	-2.246571E-02	2.206692E-02	3.987911E-04
PetalWidth	-5.747273E-03	-4.943066E-02	5.517793E-02

Classification Count Table for Iris

Actual	Predicted			Total
	Setosa	Versicolor	Virginica	
Setosa	50	0	0	50
Versicolor	0	48	2	50
Virginica	0	1	49	50
Total	50	49	51	150

Reduction in classification error due to X's = 97.0%

Misclassified Rows Section

Row	Actual	Predicted	Percent Chance of Each Group		
			Pcnt1	Pcnt2	Pcnt3
5	Virginica	Versicolo	0.0	72.9	27.1
9	Versicolo	Virginica	0.0	25.3	74.7
12	Versicolo	Virginica	0.0	14.3	85.7

Canonical Variate Analysis Section

Fn	Inv(W)B Eigenvalue	Ind'l Pcnt	Total Pcnt	Canon Corr	Canon Corr2	F-Value	Numer DF	Denom DF	Prob Level	Wilks' Lambda
1	32.191929	99.1	99.1	0.9848	0.9699	199.1	8.0	288.0	0.0000	0.023439
2	0.285391	0.9	100.0	0.4712	0.2220	13.8	3.0	145.0	0.0000	0.777973

The F-value tests whether this function and those below it are significant.

Canonical Coefficients

Variable	Canonical Variate	
	Variate1	Variate2
Constant	-2.105106	6.661473
SepalLength	-0.082938	-0.002410
SepalWidth	-0.153447	-0.216452
PetalLength	0.220121	0.093192
PetalWidth	0.281046	-0.283919

Canonical Variates at Group Means

Iris	Canonical Function	
	Function1	Function2
Setosa	-7.6076	-0.215133
Versicolor	1.82505	0.7278996
Virginica	5.78255	-0.5127666

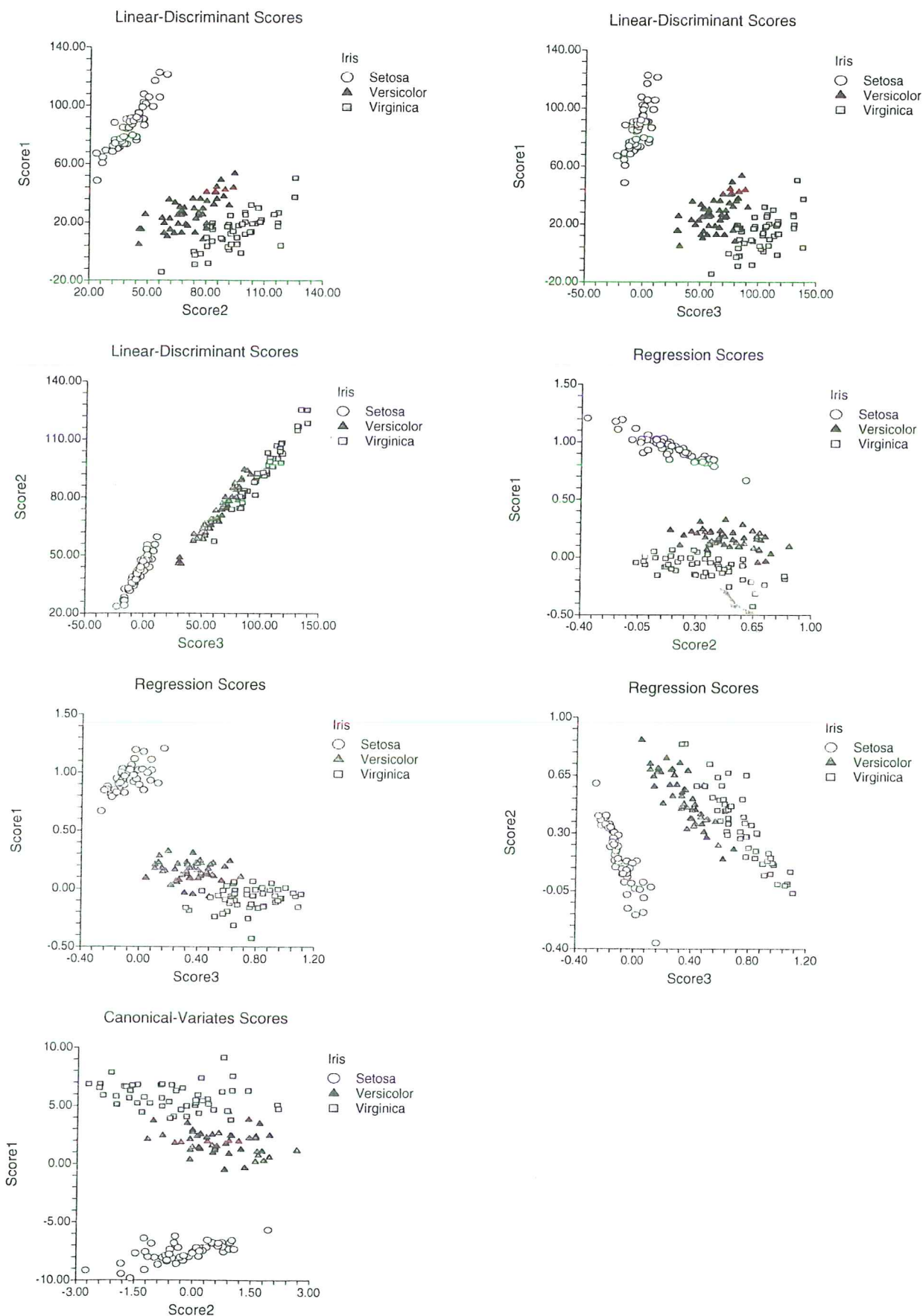
Std. Canonical Coefficients

Variable	Canonical Variate	
	Variate1	Variate2
SepalLength	-0.426955	-0.012408
SepalWidth	-0.521242	-0.735261
PetalLength	0.947257	0.401038
PetalWidth	0.575161	-0.581040

Variable-Variate Correlations

Variable	Canonical Variate	
	Variate1	Variate2
SepalLength	0.222596	-0.310812
SepalWidth	-0.119012	-0.863681
PetalLength	0.706065	-0.167701
PetalWidth	0.633178	-0.737242

Plots Section



4. Kanoninen analyysi

Seuraavilla sivuilla on esitetty Ncss-ohjelman tulostus eräästä kanonisesta analyysistä.

Aineistona on 48 parasta kymmenottelijaa vuodelta 1973. Muuttujina ovat ottelijoiden tulokset (pisteiksi muunnettuna) kymmenottelun 10:stä lajimuuttujasta (100 m, Pituushyppy, Kuula, Korkeushyppy, 400 m, 110 m:n aidat, Kiekko, Seiväs, Keihäs, 1500 m).

Kanoninen analyysi on tehty 1. päivän lajien (100 m, Pituushyppy, Kuula, Korkeushyppy, 400 m) ja 2. päivän lajien (110 m:n aidat, Kiekko, Seiväs, Keihäs, 1500 m) välillä.

- (a) Esittele kanonisen analyysin perusidea ja tavoite sekä kanonisen analyysin malli ja sen osat.
- (b) Esitä kanoniselle analyysille geometrinen tulkinta.
- (c) Kuinka monta kanonista muuttujaparia valitsisit tulostuksen perusteella? Miksi?
- (d) Esitä valitsemillesi kanonisille muuttujille tulkinnat.

Kanonisten muuttujien arvot on tallennettu muuttujiksi CAN1P1, CAN2P2, CAN1P2 ja CAN2P2.

- (e) Selitä korrelaatiot kohdassa

Correlation Report

- (f) Miten tulostuksen kuviot kohdassa

Plots Section

ja taulukko kohdassa

Correlation Report

liittyvät toisiinsa.

- (g) Mitkä ovat kanonisen korrelaation ja tavallisen korrelaation yhteydet?
- (h) Mitkä ovat kanonisen analyysin ja regressioanalyysin yhteydet?

Canonical Correlation Report

Database KYMMENOT5.S0
 Y Variables J100m, Pituush, Kuula, Korkeush, J400m
 X Variables Jaidat, Kiekko, Seivas, Keihas, J1500m

Correlation Section

	J100m	Pituush	Kuula	Korkeush	J400m
J100m	1.000000	0.171985	-0.027952	-0.411700	0.456082
Pituush	0.171985	1.000000	-0.034393	-0.003325	0.133463
Kuula	-0.027952	-0.034393	1.000000	0.162542	-0.303707
Korkeush	-0.411700	-0.003325	0.162542	1.000000	-0.338827
J400m	0.456082	0.133463	-0.303707	-0.338827	1.000000
Jaidat	0.315991	0.298065	0.086498	-0.038964	0.175548
Kiekko	0.014343	0.020875	0.727330	0.216995	-0.344647
Seivas	0.054723	0.061024	-0.204229	-0.117765	0.006584
Keihas	-0.221250	0.153750	0.023140	0.149779	-0.104689
J1500m	-0.291744	-0.206665	-0.446248	-0.146150	0.302178
	Jaidat	Kiekko	Seivas	Keihas	J1500m
J100m	0.315991	0.014343	0.054723	-0.221250	-0.291744
Pituush	0.298065	0.020875	0.061024	0.153750	-0.206665
Kuula	0.086498	0.727330	-0.204229	0.023140	-0.446248
Korkeush	-0.038964	0.216995	-0.117765	0.149779	-0.146150
J400m	0.175548	-0.344647	0.006584	-0.104689	0.302178
Jaidat	1.000000	0.047707	-0.073497	-0.148188	-0.224552
Kiekko	0.047707	1.000000	-0.181800	0.135561	-0.573501
Seivas	-0.073497	-0.181800	1.000000	-0.128523	0.012496
Keihas	-0.148188	0.135561	-0.128523	1.000000	-0.065414
J1500m	-0.224552	-0.573501	0.012496	-0.065414	1.000000

Canonical Correlations Section

Variate Number	Canonical Correlation	R-Squared	F-Value	Num DF	Den DF	Prob Level	Wilks' Lambda
1	0.789434	0.623206	2.87	25	143	0.000047	0.220523
2	0.528435	0.279244	1.43	16	120	0.136882	0.585261

F-value tests whether this canonical correlation and those following are zero.

Variation Explained Section

Canonical Variate Number	Variation in these Variables	Explained by these Variates	Individual Percent Explained	Cumulative Percent Explained	Canonical Correlation Squared
1	Y	Y	22.0	22.0	0.6232
2	Y	Y	21.4	43.4	0.2792
1	Y	X	13.7	13.7	0.6232
2	Y	X	6.0	19.7	0.2792
1	X	Y	19.9	19.9	0.6232
2	X	Y	4.3	24.3	0.2792
1	X	X	32.0	32.0	0.6232
2	X	X	15.5	47.5	0.2792

Standardized Y Canonical Coefficients Section

	Y1	Y2
J100m	0.458881	0.873347
Pituush	0.118940	0.450449
Kuula	0.771738	-0.500897
Korkeush	0.195456	0.142655
J400m	-0.348227	-0.388564

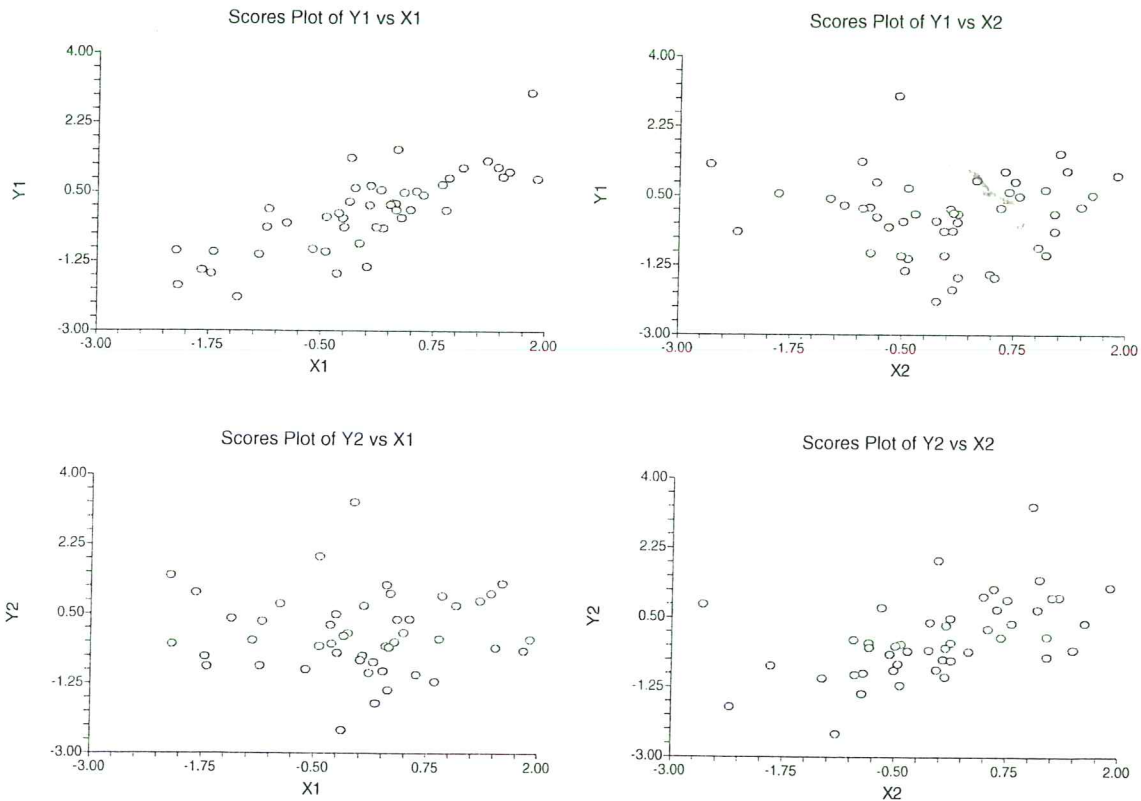
Standardized X Canonical Coefficients Section

	X1	X2
Jaidat	0.082409	0.410385
Kiekko	0.703672	-0.826217
Seivas	-0.066863	0.190447
Keihas	-0.116921	0.000540
J1500m	-0.391197	-0.881017

Variable - Variate Correlations Section

	Y1	Y2	X1	X2
J100m	0.218476	0.728871	0.172472	0.385161
Pituush	0.124193	0.565546	0.098042	0.298855
Kuula	0.892349	-0.399604	0.704451	-0.211165
Korkeush	0.249568	-0.168160	0.197017	-0.088862
J400m	-0.423674	0.173663	-0.334463	0.091770
Jaidat	0.178462	0.293137	0.226063	0.554726
Kiekko	0.732801	-0.177515	0.928261	-0.335925
Seivas	-0.150552	0.158220	-0.190709	0.299412
Keihas	0.000348	-0.073517	0.000441	-0.139123
J1500m	-0.636636	-0.262626	-0.806446	-0.496988

Plots Section



Correlation Report

Database KYMMENOT5.S0

Pearson Correlations Section (Row-Wise Deletion)

	Can1P1	Can2P1	Can1P2	Can2P2
Can1P1	1.000000	0.789434	0.000000	0.000000
Can2P1	0.789434	1.000000	0.000000	0.000000
Can1P2	0.000000	0.000000	1.000000	0.528435
Can2P2	0.000000	0.000000	0.528435	1.000000

Cronbachs Alpha = 0.529605 Standardized Cronbachs Alpha = 0.529605

5. Hierarkkinen ryhmittelyanalyysi

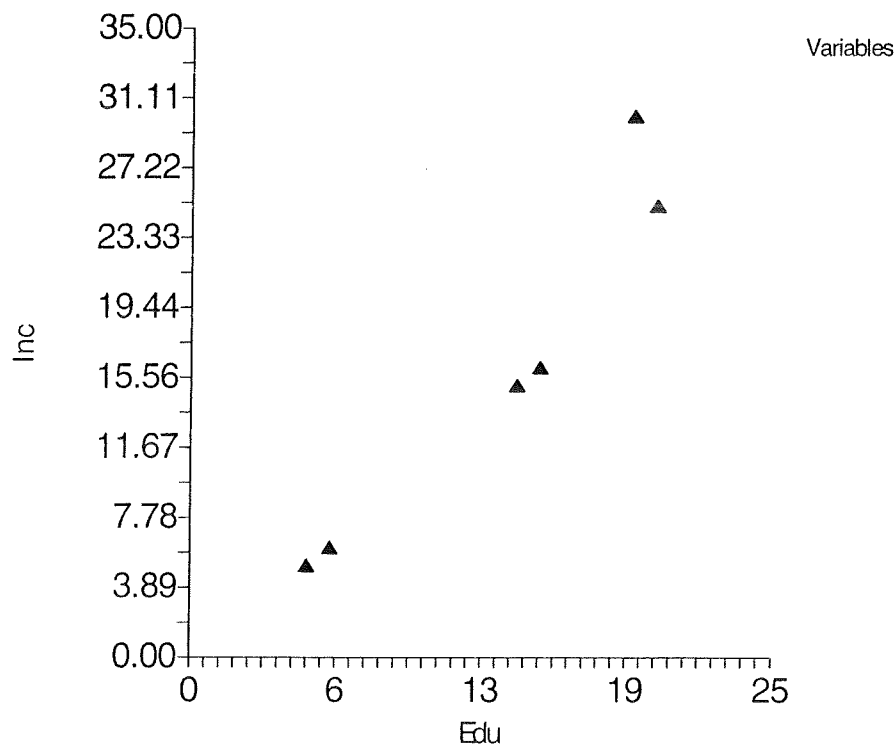
Seuraavilla sivuilla on esitetty Ncss-ohjelman tulostus eräästä ryhmittelyanalyysistä.

Analysoitavana on tulojen ja koulutuksen pituuden riippuvuutta kuvaava pieni havaintoaineisto. Muuttujina ovat

Edu = koulutuksen pituus vuosina

Inc = vuositulot (1000 \$)

Scatter Plot Section



- Esittele hierarkkisen ryhmittelyanalyysin perusidea ja tavoite.
- Selitä miten tulostuksessa sovellettu ryhmittelymenetelmä (mikä menetelmä?) toimii.
- Konstruoi käytetyn etäisyysmitan mukainen etäisyysmatriisi havaintopisteille.
- Mitä tarkoitetaan tulostuksessa mainitulla kofeneettisella korrelaatiolla?
- Kuvaa tulostuksen dendrogrammin idea.
- Esittele lyhyesti muita hierarkkisen ryhmittelyn menetelmiä.

Hierarchical Clustering Report

Database CLUSTDIST1.S0
 Variables Edu, Inc
 Clustering Method Centroid (Unweighted Pair-Group Centroid)
 Distance Type Euclidean
 Scale Type None

Cluster Detail Section

Row	Cluster
1	1
2	1
3	2
4	2
5	3
6	3

Linkage Section

Link	Number Clusters	Distance Value	Distance Bar	Rows Linked
5	1	194.125000		1,2,3,4,5,6
4	2	84.500000		3,4,5,6
3	3	13.000000		5,6
2	4	1.000000		3,4
1	5	1.000000		1,2

Cophenetic Correlation 0.665827
 Delta(0.5) 0.587618
 Delta(1.0) 0.665843

Distance Section

First Row	Second Row	Actual Distance	Dendrogram Distance	Actual Difference	Percent Difference
1	2	1.000000	1.000000	0.000000	0.00
1	3	90.500000	194.125000	-103.625000	-114.50
1	4	110.500000	194.125000	-83.625000	-75.68
1	5	312.500000	194.125000	118.375000	37.88
1	6	410.500000	194.125000	216.375000	52.71
2	3	72.500000	194.125000	-121.625000	-167.76
2	4	90.500000	194.125000	-103.625000	-114.50
2	5	278.500000	194.125000	84.375000	30.30
2	6	372.500000	194.125000	178.375000	47.89
3	4	1.000000	1.000000	0.000000	0.00
3	5	68.000000	84.500000	-16.500000	-24.26
3	6	125.000000	84.500000	40.500000	32.40
4	5	53.000000	84.500000	-31.500000	-59.43
4	6	106.000000	84.500000	21.500000	20.28
5	6	13.000000	13.000000	0.000000	0.00

Dendrogram Section

Dendrogram

