Παραγοντική Ανάλυση Αντιστοιχιών και Ανιούσα Ιεραρχική Ταξινόμηση με το JAMOVI

1. Εισαγωγή Πίνακα Δεδομένων στο JAMOVI

Στο πρώτο παράθυρο που εμφανίζεται με την εκτέλεση του προγράμματος επιλέγουμε

και Open (This PC) → Browse (Εικόνα 1). Εντοπίζουμε το αρχείο δεδομένων μας (π.χ. αρχείο .csv) και επιλέγουμε Open. Προσοχή: η πρώτη γραμμή του αρχείου πρέπει να περιέχει τα ονόματα των μεταβλητών και η πρώτη στήλη τα ονόματα των αντικειμένων/υποκειμένων.

						Un	titled	
= 🖊	Variab!es	Data	Analyses	Edit				
Exploration	T-Tests	₽ ₽ ANOA	Regression	Frequencies	Factor	snowCluster	SEM	MEDA 2
🐣 A	🐣 E	•••				Untitled		
1 2 3		÷	jamovi	↑ Home			Browse	
4 5		New		180-Artic	le Text-1020-2-2-	-20220411 2-20220411		
6		Open		19617396	3			
8		This PC Data Librar	у	2				
9		Special Imp	port	5η ΟΣΣ				
11		Save		pycach	ne			
12								

Εικόνα 1. Εισαγωγή αρχείου δεδομένων στο JAMOVI

Τα δεδομένα έχουν εισαχθεί στον πίνακα που βρίσκεται κάτω από την γραμμή εργαλείων (βλ. Εικόνα 2). Στο παράδειγμά μας, πρόκειται για 21,187 υποκείμενα (άτομα) και 16 μεταβλητές (ερωτήσεις σχετικά με τις διακρίσεις που πιθανόν υφίστανται τα άτομα). Μπορούμε να αποθηκεύσουμε τα δεδομένα σε αρχείο του

JAMOVI (με επέκταση .omv) με επιλογή του 🧮 και Save/Save as...

		s Data	Analyses	Edit					≤: :
Ex	ploration T-Test	ts ANOVA	Regression Free	quencies Factor	Flexplot	snowRMM	snowCluster		θ Modules Γραμμή
	😪 cntry	🐣 icgndra	🐣 pdjobev	🐣 dscrgrp	🐣 dscrrce	👌 dscrntn			- prophetican
3	AT	6	6	2	0				εργαλειων
4	AT	6	6	2	0				
5	AT	6	6	2	0				
6	AT	6	6	2	0				
7	AT	6	6	2	0				
8	AT	6	6	2	0				
9	AT	6	6	2	0				
10	AT	6	6	2	0				
11	AT	6	6	2	0			version 2.3.28	
12	AT	6	6	2	0				
13	AT	6	6	2	0				
14	AT	6	6	2	0				
15	AT	6	6	2	0				
16	AT	6	6	2	0				
F 11	cáng 2 T	a SeSoué		TAMO	VI rear n	NORILLI	n courd cion		

Εικόνα 2. Τα δεδομένα μας στο JAMOVI και η γραμμή εργαλείων

Ελέγχουμε ότι τα δεδομένα έχουν εισαχθεί σωστά στην καρτέλα Data. Παρατηρήστε ότι το JAMOVI έχει αναγνωρίσει αυτόματα την κλίμακα μέτρησης των μεταβλητών ως Nominal (κατηγορική μεταβλητή σε ονομαστική κλίμακα). Με διπλό κλικ στο όνομα μιας μεταβλητής μπορούμε να δούμε περισσότερες πληροφορίες γι' αυτήν, όπως την κλίμακα μέτρησής της (Measure type) (Εικόνα 3).

≡	Variables	Data	Analyses	Edit						
Exploration	T-Tests	<mark>ф</mark> ANOVA	Regression	Frequencies	Factor	Flexplot	snowRMM	snowCluster		
			DATA VARIA	BLE						
			icgndra							U
			Description							
			Measure typ	e Nominal	▼ 🐣		Levels		\uparrow	
		/	Data type 🛛	nteger 🗸		1			\checkmark	
			Missing valu	es		2				
						6				
						Retain u	nused levels	in analyses 🔘		

Εικόνα 3. Κλίμακα μέτρησης μιας μεταβλητής (Measure type)

Αν η κλίμακα μέτρησης δεν έχει αναγνωριστεί σωστά από το λογισμικό τότε χρειάζεται να την ορίσουμε εμείς.

Προκειμένου να εφαρμόσουμε την Παραγοντική Ανάλυση των Αντιστοιχιών (Multiple Correspondence Analysis), πρέπει να αλλάζουμε, για κάθε μεταβλητή, το πεδίο Data type από Integer σε Text, όπως στην Εικόνα 4. Επαναλαμβάνουμε το ίδιο για όλες τις μεταβλητές που πρόκειται να αναλύσουμε με τη συγκεκριμένη μέθοδο.

≡	Variables	Data	Analyses	Edit						
Exploration	₽ T-Tests	<mark>ұ</mark> ұ ANOVA	Regression	Frequencies	Factor	Flexplot	snowRMM	snowCluster		
			DATA VARIA	BLE						
			Description							
			Measure type	e Nominal	Y S		Levels		\uparrow	
			Data type 🔳	ext 🗸		1			\checkmark	>
			Missing value	es		2				
						6				
									+	
						Retain u	nused levels	in analyses 🔾		

Εικόνα 4. Αλλαγή τύπου δεδομένων σε 'Text'

Επειδή το αρχείο δεδομένων περιλαμβάνει τα άτομα από κάθε χώρα που συμμετείχαν στην έρευνα EVS, πρέπει να επιλέξουμε μία συγκεκριμένη χώρα. Για παράδειγμα, θα επιλέξουμε την Γερμανία, η οποία αντιστοιχεί στην τιμή "DE" της μεταβλητής entry. Στην καρτέλα Data επιλέγουμε Filters. Στο πεδίο *fx* πληκτρολογούμε cntry="DE", όπως στην παρακάτω Εικόνα.

0 •				Discrimina	ation		
😑 Variables	Data	Analyses	Edit				
Paste □ Clipboard Edit	Setup	Compute	Transform Variables	Add × Delete	Filters	Add × Delete Rows	
	٦ • •		S r 1 = <u>cntry=="C</u> cription	E		active X +	>

Εικόνα 5. Φιλτράρισμα των υποκειμένων (ατόμων) που αντιστοιχούν σε μια συγκεκριμένη χώρα

2. Εφαρμογή της Παραγοντικής Ανάλυσης των Αντιστοιχιών (Multiple Correspondence Analysis)

Για να εφαρμόσουμε την Πολλαπλή Παραγοντική Ανάλυση Αντιστοιχιών (Multiple Correspondence Analysis - MCA) χρειάζεται να έχουμε εγκαταστήσει το πρόσθετο (module) snowCluster.

Κλικ στο → jamovi library, εντοπίζουμε το πρόσθετο snowCluster και κλικ

στο INSTALL (Εικόνα 6).

	Installed	Available	Sideload
Qsnowcluster			
	snowCluster - Multivari Hyunsoo Seol This module allows users to analyze Multiple Factor Analysis, Factor ana Scaling, Univariate time series, Prop visualization results.	ate Analys k-means and h lysis of mixed d whet analysis, De	is 7.2.8 ierarchical clustering,Correspondence Analysis, ata, Discriminant Analysis, Multidimensional ecision Tree, Machine Learning,and various

Εικόνα 6. Εγκατάσταση module snowCluster από την βιβλιοθήκη του JAMOVI

Βεβαιωνόμαστε ότι το πρόσθετο έχει εγκατασταθεί και εμφανίζεται στη γραμμή εργαλείων (καρτέλα Analyses).

Κλικ στο snowCluster από τη γραμμή εργαλείων και στη συνέχεια επιλέγουμε Multiple Correspondence Analysis. Μεταφέρουμε τις μεταβλητές στη λίστα Variables (Εικόνα 7).

Cluster				
K-means Clustering				
Hierarchical Clustering				
Clustering Dendrogram				
Multidimensional Scaling Plot				
PCA Plot	A definition of the second	A		\bigcirc
Group plot	viultiple Correspondence	Anaiysi	S	
Correspondence & Factor				<u> </u>
Correspondence Analysis			Variables	
Multiple Correspondence Analysis	Ca cntry			
Multiple Factor Analysis		7		
Factor analysis of mixed data			🛃 pdjobev	
Classification			🔗 dscrgrp	
Linear Discriminant Analysis				
Decision Tree			🐣 dscrntn	
Machine Learning			Factor(Optional)	
		\rightarrow		

Εικόνα 7. Η επιλογή μεταβλητών για την ΜCA

Από τις επιλογές που βρίσκονται στο κάτω μέρος (κλικ στο Multiple Correspondence Analysis) επιλέγουμε Variable categories με Type cos2 και Individuals με Type coordinates. Από τα Plots, επιλέγουμε Biplot και Coordinates of variables categories όπως στην Εικόνα 8.

 Multiple Correspondence Analysis

 Multiple Correspondence Analysis
 Eigenvalues
 Variable categories
 Type cos2

 Individuals

 Type coordinates

 Plots

 Scree plot
 Biplot
 Correlation between variables and dimensions
 Coordinates of variable categories
 Plot of individuals

Στο δεξί μέρος της οθόνης εμφανίζονται τα αποτελέσματα, οι τιμές του δείκτη cos2 (COR) των στηλών (κατηγορίες των μεταβλητών - variable categories) και των γραμμών (άτομα - individuals) για τους πέντε πρώτους παραγοντικούς άζονες, καθώς και το παραγοντικό επίπεδο (biplot) με τα σημεία γραμμών και στηλών (όπως και το παραγοντικό επίπεδο μόνο με τα σημεία στηλών).

Αντιγράφουμε στο αρχείο κειμένου της εργασίας μας τον πίνακα "Variable categories". Επιλέγουμε ένα όριο για τον δείκτη ποιότητας προβολής COR ή cos2 (π.χ. το 0,2), και σημειώνουμε τα σημεία με δείκτη COR > 0,2 στην πρώτη ή τη δεύτερη στήλη του πίνακα. Αυτά είναι τα σημεία που θα ερμηνεύσουμε στο παραγοντικό επίπεδο (βλ πρότυπο εργασίας).

Μπορούμε να την επιβεβαιώσουμε ή να την αλλάξουμε/εμπλουτίσουμε αυτά τα αποτελέσματα με βάση τα αποτελέσματα της Ανιούσας Ιεραρχικής Ταξινόμησης στις παραγοντικές συντεταγμένες (βλ. επόμενη ενότητα).

Αντιγραφή του πίνακα συντεταγμένων των ατόμων στα αρχικά δεδομένα

Πριν εφαρμόσουμε την Ανιούσα Ιεραρχική Ταξινόμηση, πρέπει να μεταφέρουμε στα δεδομένα των πίνακα με τις συντεταγμένες των ατόμων στους πέντε πρώτους παραγοντικούς άξονες. Αυτές οι τιμές βρίσκονται στον πίνακα Individuals across dimensions – coordinates.

Από το παράθυρο με τα αποτελέσματα, αντιγράφουμε με δεξί κλικ και Copy τον πίνακα αυτόν. Κάνουμε επικόλληση (paste) σε ένα κενό αρχείο Excel και σβήνουμε τα περιττά ώστε να μην υπάρχουν ενδιάμεσες κενές στήλες. Μεταβαίνουμε στην καρτέλα Data του Jamovi, και επιλέγουμε το πρώτο κελί που βρίσκεται δίπλα στην τελευταία στήλημεταβλητή με τα δεδομένα μας (προσοχή: το κελί αυτό βρίσκεται στη γραμμή που βρίσκεται και το πρώτο άτομο της χώρας που επιλέζαμε να αναλύσουμε). Κάνουμε

Εικόνα 8. Οι επιλογές της MCA

εκεί copy-paste από το Excel μόνο τις τιμές (όχι δηλαδή τα ονόματα των μεταβλητών Dim1, Dim2 ...), όπως φαίνεται στην Εικόνα 9. Αρχικά τα ονόματα των μεταβλητών θα είναι γράμματα του αγγλικού αλφαβήτου (π.χ. K, L, M, N, O), αλλά μπορούμε να τα μετονομάσουμε με διπλό κλικ στο όνομα της κάθε στήλης σε Dim1, Dim2, Dim3, Dim4 και Dim 5.

Τελικά, έχουμε δημιουργήσει τις πέντε νέες στήλες που αντιστοιχούν στις συντεταγμένες των ατόμων στους πέντε πρώτους παραγοντικούς άξονες (Εικόνα 9).

≡		Variables	Data	Analys	es Edit					
Paste	oard	 	Setup	Comput	te Transform Variables	Add ~	Filters	Add ~ Delete		
dscrdsb		臱 dscroth	臱 ctzo	ntr	😪 brncntr	🤌 Dim1	🤣 Dim2	🔶 Dim3	🔶 Dim4	march
/401	1	0		1	1					
7482	1			1	2					Variable cat
7483	0	1		1	1					
7484	0	1		1	Conv-past	ε εδώ				
7485	0	1		1	copy pus					
7486	0	1		1	1					
7487				1	1	7				
7488	0	0		1	1	-0.198	-0.082	-0.082	0.0	Individuals a
7489	0	0		1	1	-0.198	-0.082	-0.082	0.0	
7490	0	0		1	1	-0.198	-0.082	-0.082	0.0	
7491	0	0		1	1	-0.198	-0.082	-0.082	0.0	
7492	0	0		1	1	-0.198	-0.082	-0.082	0.0	
7493	0	0		1	1	-0.198	-0.082	-0.082	0.0	
7494	0	0		1	1	-0.198	-0.082	-0.082	0.0	Biplot
7495	0	0		1	1	-0.198	-0.082	-0.082	0.0	Ĩ
7496	0	0		1	1	-0.198	-0.082	-0.082	0.0	
7497	0	0		1	1	-0.198	-0.082	-0.082	0.0	
7498	0	0		1	1	-0.198	-0.082	-0.082	0.0	

Εικόνα 9. Αντιγραφή των παραγοντικών συντεταγμένων των ατόμων στους 5 άξονες, ως νέες μεταβλητές

4. Εφαρμογή της Ανιούσας Ιεραρχικής Ταξινόμησης στις παραγοντικές συντεταγμένες

Πηγαίνουμε στο Analyses → snowCluster από τη γραμμή εργαλείων και στη συνέχεια Clustering Dendrogram. Μεταφέρουμε τις πέντε νέες στήλες (Dim1 έως Dim5) στη λίστα Variables (Εικόνα 10).

≡	≡ ۱	/ariables	Data Analyses Edit			: ا
Exp	loration	₽ T-Tests	ANOVA Regression Frequencies F	actor Flexplot snowRMM snowCluster		Modules
1	Filter 1	AT AT	Clustering Dendrogram	(ightarrow		Clustering Dendrog
3 4	××	AT AT	🐣 dscrage	Q Variables		Instructions
5 6	××	AT AT	e decreax	→ Pim5 → Dim4		1. Do not move any variable into
7 8	××	AT AT	decret	Dim3 Dim2		 Cluster numbers will be displi Feature requests and bug reg
9 10	××	AT		Labels (Optional)		
11 12 12	××	AT	Clustering		0	
13 14 15	×	AT	✓ Hierarchical Clustering			
16	×	AT	Options	Plot		
17	×	AT	Number of clusters 3	Cluster Dendrögräm		References
19 20	×	AT	Distance measure euclidean V			[1] The jamovi project (2022). <i>jamovi.</i> (Version 2-3)
21 Rea	. ▼R(<	Fil DFilteAc Ce	Clustering method ward.D2 V			[Computer Software].

Εικόνα 10. Η επιλογή των μεταβλητών για την Ανιούσα Ιεραρχική Ταξινόμηση

Στη συνέχεια αφαιρούμε τις επιλογές Standardize data και Cluster dendrogram. Στο πεδίο Number of Clusters βάζουμε τον αριθμό των ομάδων που εντοπίσαμε προηγουμένως στο παραγοντικό επίπεδο (έστω 3). Επιπλέον, επιλέγουμε Save Cluster number (για την αποθήκευση της μεταβλητής ομαδοποίησης – νέα στήλη Clustering – στα δεδομένα), όπως στην Εικόνα 11.

Ελέγχουμε ότι στην καρτέλα Data (δλδ. στα δεδομένα μας) έχει δημιουργηθεί στο τέλος μια νέα στήλη, η μεταβλητή ομαδοποίησης – νέα στήλη Clustering (Εικόνα 12).

✓ Hierarchical Clustering	
Options	Plot
Standardize data	Cluster Dendrogram
Number of clusters 3	
Distance measure euclidean V	
Clustering method ward.D2 V	
Type of plot rectangle V	
✓ Save	
Cluster number	

Εικόνα 11. Οι επιλογές της Ανιούσας Ιεραρχικής Ταξινόμησης

≡	Va	riables	Data	Analyses	e Edit	
Past ^{Clip}	e Doard	Sedit	Setup	Compute	Transform Variables	
	🧼 Dim	4	🔶 Dim5	🐣 Clu	istering	•
7473						
7474						
7475						
7476						
7477						
7478						
7479						
7480						
7481						
7482						
7483						
7484						
7485						
7486						
7487						
7488		0.017	0.0)04		1
7489		0.017	0.0	004		1
7490		0.017	0.0	004		1
7491		0.017	0.0	004		1

Εικόνα 12. Η νέα στήλη Clustering

5. Ερμηνεία των ομάδων με βάση τις αρχικές μεταβλητές

Για να ερμηνεύσουμε τις ομάδες μπορούμε να τις περιγράψουμε ως προς τις αρχικές μεταβλητές. Προηγουμένως, όμως, <u>πρέπει να μετατρέψουμε όλες τις αρχικές μεταβλητές σε ποσοτικές (continuous).</u> Για να το κάνουμε αυτό πηγαίνουμε στα Data και κάνουμε διπλό κλικ στο όνομα της κάθε μεταβλητής. Αλλάζουμε το Measure Type σε Continuous, βλ. Εικόνα 13 για μία μεταβλητή. Αυτό πρέπει να γίνει για όλες τις αρχικές μεταβλητές.

≡	Variables	Data	Analyses	Edit				
Paste Clipboard	► C	Setup	Compute	Transform Variables	Add	Filters	Add ~ Delete Rows	
			DATA VARI	ABLE				
			Description	1				
			Measure ty	e Continuo	us 🗸 🗸		Levels	1
			Data type	Decimal 🗸				1
			Missing valu	les				
						Retain un	used levels in analyse	s 💽

Εικόνα 13. Μετατροπή μεταβλητής σε Continuous

Επιλέγουμε από τη γραμμή εργαλείων Exploration → Descriptives. Στη λίστα Variables μεταφέρουμε τις αρχικές μεταβλητές και στο πεδίο Split μεταφέρουμε τη μεταβλητή ομαδοποίησης (Clustering), όπως φαίνεται παρακάτω.

≡	Variables	Data	Analyses	Edit										<u>.</u>
Exploratio	on T-Tests	₽ ANOA	Regression	Frequencies	Factor	Flex	plot	snowRMM	snowCluste	r			М	dules
🔶 ds 7483	scrgrp 🤌	Desc	riptives							()		Descri	ptives	
7484	1	Dese	npuves							\bigcirc				
7485	1											Descriptive	s	
7486	1	🔒 🛃 cnt	try		Q	L L	ariable	S					Clusterin	g
7487	1	icg	ndra			\rightarrow	🧳 dso	crgrp				dscrarp	1	1
7488	2	ba 🔶	iobev				🧼 dso	crrce				asergip	2	
7489	2		n1				🤶 dso	orntn					3	
7490	2						< dso	orrig				dscrrce	1	1
7491	2		112				ieh 🚫	rina					2	
7492	2	I Dir	n3			S	Split by						3	
7493	2	n 🔶 Dir	n4			\rightarrow	🐣 Clu	Istering			:	dscrntn	1	1
7494	2	ir 🥠 Dir	n5										2	
7495	2												3	
7496	2	Descrip	tives Variable	es across row	s v			Er Fr	equency tab	les 🐣 📲		dscrrlg	1	1
7497	2												2	
7498	2	🗸 S	tatistics										3	
7499	2		~									dscring	1	1
7500	2	Sample	Size				Cent	ral lendend	сy				2	
7501	2	N	Missing				 ✓ 	Mean					3	
7502	2							Median				dscretn	1	1
7502	2	Percen	tile values					wealdri					2	

Εικόνα 14. Περιγραφή των ομάδων με βάση τις αρχικές μεταβλητές

Στο παράθυρο αυτό επιλέγουμε Descriptives Variables across rows και από τα Statistics, μόνο N και Mean.

Αν κοιτάξουμε τα αποτελέσματα (Εικόνα 15) και εστιάσουμε στους μέσους όρους (Mean) θα διαπιστώσουμε ότι τα άτομα στις ομάδες 1 (1243 άτομα, 79,17%) και 2 (103 άτομα, 6,5%) περιγράφουν τον εαυτό τους ως άτομα που δεν υφίστανται διακρίσεις (μέσοι όροι 1,99 και 1,78 – θυμίζουμε εδώ ότι στη συγκεκριμένη μεταβλητή το 2

σημαίνει Όχι και το 1 = Ναι), ενώ η ομάδα 3 περιλαμβάνει άτομα που υφίστανται διακρίσεις (224 άτομα, 14,26%), με μέσο όρο 1.

Descri	ptives		
Descriptive	S		
	Clustering	Ν	Mean
dscrgrp	1	1243	1.99678
	2	103	1.78641
	3	224	1.00000
dscrrce	1	1243	0.00000
	2	103	0.07767
	3	224	0.23661
dscrntn	1	1243	0.00000
	2	103	0.08738
	3	224	0.08482
dscrrlg	1	1243	0.00000
	2	103	0.02913
	3	224	0.18750
dscring	1	1243	0.00000
	2	103	0.00000
	3	224	0.03571
dscretn	1	1243	0.00000
	2	103	0.00000
	3	224	0.16071

Εικόνα	15.	Μέσοι	όροι	των τ	οιών	ομάδων	στις	αργικές	μεταβλη	ιτές
		1110000	~ ~ ~ ·		P	0 0000000	~ ~ ~ ~		Pro . 01 p. 11	
							2			