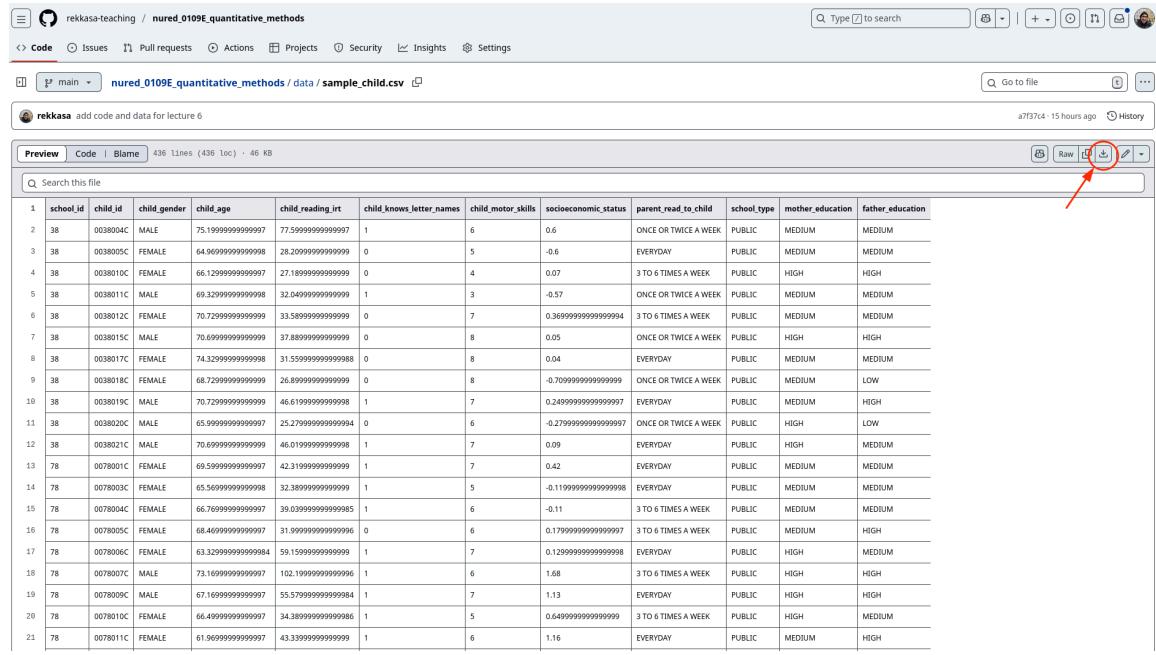


# Εισαγωγή και επεξεργασία δεδομένων

## Εισαγωγή δεδομένων

Για τις ανάγκες της ανάλυσης θα χρησιμοποιήσουμε τα δεδομένα που βρίσκονται εδώ. Για να τα κατεβάσετε απλά πατάτε το κουμπί download στο πάνω δεξιά μέρος της σελίδας.

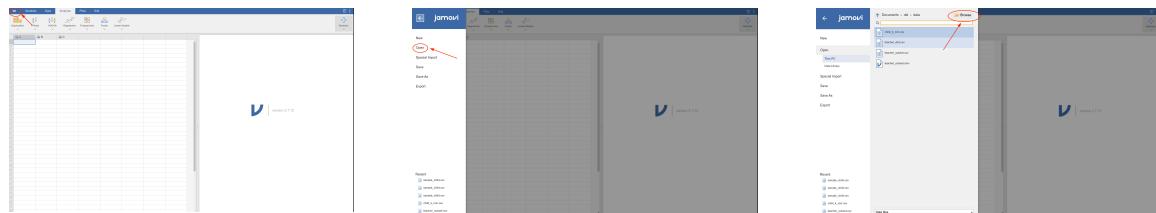


school_id	child_id	child_gender	child_age	child_reading_irt	child_knows_letter_names	child_motor_skills	socioeconomic_status	parent_read_to_child	school_type	mother_education	father_education
1	38	0038004C	MALE	75.19999999999997	77.59999999999997	1	6	0.6	ONCE OR TWICE A WEEK	PUBLIC	MEDIUM
2	38	0038005C	FEMALE	64.96999999999998	28.20999999999998	0	5	-0.6	EVERYDAY	PUBLIC	MEDIUM
3	38	0038010C	FEMALE	66.12999999999997	27.18999999999997	0	4	0.07	3 TO 4 TIMES A WEEK	PUBLIC	HIGH
4	38	0038011C	MALE	69.32999999999998	32.04999999999999	1	3	-0.57	ONCE OR TWICE A WEEK	PUBLIC	MEDIUM
5	38	0038012C	FEMALE	70.72999999999999	33.58999999999999	0	7	0.3699999999999994	3 TO 6 TIMES A WEEK	PUBLIC	MEDIUM
6	38	0038017C	MALE	70.69999999999999	37.88999999999999	0	8	0.05	ONCE OR TWICE A WEEK	PUBLIC	HIGH
7	38	0038017C	FEMALE	74.32999999999998	31.55999999999998	0	8	0.04	EVERYDAY	PUBLIC	MEDIUM
8	38	0038019C	FEMALE	68.72999999999999	26.89999999999999	0	8	-0.709999999999999	ONCE OR TWICE A WEEK	PUBLIC	MEDIUM
9	38	0038020C	MALE	70.72999999999999	46.61999999999998	1	7	0.2499999999999997	EVERYDAY	PUBLIC	HIGH
10	38	0038021C	MALE	65.99999999999997	25.27999999999994	0	6	-0.2799999999999997	ONCE OR TWICE A WEEK	PUBLIC	HIGH
11	38	0038021C	MALE	70.69999999999999	46.01999999999998	1	7	0.09	EVERYDAY	PUBLIC	HIGH
12	38	0038021C	FEMALE	69.59999999999997	42.31999999999998	1	7	0.42	EVERYDAY	PUBLIC	MEDIUM
13	78	0078003C	MALE	65.56999999999998	32.38999999999998	1	5	-0.1199999999999998	EVERYDAY	PUBLIC	MEDIUM
14	78	0078004C	FEMALE	66.76999999999997	39.03999999999985	1	6	-0.11	3 TO 4 TIMES A WEEK	PUBLIC	MEDIUM
15	78	0078005C	FEMALE	68.46999999999997	31.09999999999996	0	6	0.1799999999999997	3 TO 6 TIMES A WEEK	PUBLIC	HIGH
16	78	0078006C	MALE	63.32999999999984	59.15999999999999	1	7	0.1299999999999998	EVERYDAY	PUBLIC	HIGH
17	78	0078007C	MALE	73.16999999999997	102.19999999999996	1	6	1.68	3 TO 6 TIMES A WEEK	PUBLIC	HIGH
18	78	0078009C	MALE	67.16999999999997	55.57999999999984	1	7	1.13	EVERYDAY	PUBLIC	HIGH
19	78	0078010C	FEMALE	66.49999999999997	34.38999999999986	1	5	0.6499999999999998	3 TO 6 TIMES A WEEK	PUBLIC	HIGH
20	78	0078011C	FEMALE	61.96999999999997	43.33999999999991	1	6	1.16	EVERYDAY	PUBLIC	MEDIUM
21	78	0078011C	FEMALE	61.96999999999997	43.33999999999991	1	6	1.16	EVERYDAY	PUBLIC	HIGH

Figure 1: Κουμπί download

Το αρχείο που θα κατέβει ονομάζεται `sample_child.csv`. Αρχεία αυτού του τύπου (csv) μπορείτε να τα ανοίξετε με το excel ή ακόμα και με ένα πρόγραμμα όπως το σημειωματάριο. Τα αρχεία αυτά χρησιμοποιούν κόμματα για να ορίσουν μέτρηση για μια νέα μεταβλητή.

Αν έχετε ανοιχτό το πρόγραμμα Jamovi, μπορείτε να εισάγετε τα δεδομένα στο που κατεβάσατε ακολυθώντας τις εικόνες παρακάτω:



(a) Βήμα 1

(b) Βήμα 2

(c) Βήμα 3

Figure 2: Διαδικασία φόρτωσης δεδομένων στο Jamovi.

Αν όλα πάνε καλά θα έχετε τα δεδομένα σας μέσα στο Jamovi, όπως φαίνεται στο Figure 3.

The screenshot shows the Jamovi software interface. At the top, there is a navigation bar with tabs for 'Variables', 'Data', 'Analyses', 'Plots', and 'Edit'. Below the navigation bar are several icons representing different statistical analyses: Exploration (bar chart), T-Tests (t-test icon), ANOVA (ANOVA icon), Regression (regression line icon), Frequencies (frequency distribution icon), Factor (factor analysis icon), and Linear Models (linear regression icon). On the right side of the interface, there is a 'Modules' section with a plus sign icon. The main area displays a data table with columns labeled 'school\_id', 'child\_id', 'child\_ge...', 'child\_ag...', 'child\_rea...', 'child\_kno...', 'child\_mo...', and 'socieec'. The data consists of 435 rows, each containing values for these variables. The bottom of the screen shows status information: 'Row count 435', 'Filtered 0', 'Deleted 0', 'Added 0', and 'Cells edited 0'. The version of the software is indicated as 'version 2.7.12'.

Figure 3: Τα δεδομένα έχουν φορτωθεί στο Jamovi με επιτυχία.

## Αρχική επεξεργασία

Όταν εισάγουμε τα δεδομένα μας, το Jamovi κάνει ό,τι μπορεί για να προβλέψει τον τύπο των μεταβλητών. Δυστυχώς, δεν κάνει πάντα την καλύτερη δουλειά. Γι' αυτό είναι σημαντικό, μετά από την εισαγωγή των δεδομένων, πάντα να κάνετε έναν έλεγχο στον τύπο των μεταβλητών που φορτώθηκαν. Το Jamovi χρησιμοποιεί 4 είδη μεταβλητών:

1. Αναγνωριστικές μεταβλητές (ID). Αυτές χρησιμοποιούνται για να δώσουν έναν αναγνωριστικό (συνήθως) αριθμό σε κάθε π.χ. άτομο, σχολείο κτλ. Τις περισσότερες φορές, τέτοιες μεταβλητές δεν παίζουν και μεγάλο ρόλο στις αναλύσεις μας. Στο Jamovi αυτές συμβολίζονται με ένα μικρό ταμπλέάκι.
2. Ονομαστικές μεταβλητές (Nominal variables). Αυτές οι μεταβλητές μας δίνουν κατηγορίες οι οποίες δεν μπορούν να διαταχθούν. Για παράδειγμα το φύλο ή το χρώμα της τσάντας των μαθητών. Το μπλε δεν μπορούμε να πούμε ότι είναι «μεγαλύτερο» ή «μικρότερο» από το κόκκινο. Στο Jamovi αυτές συμβολίζονται με τρεις τεμνόμενους κύκλους.
3. Διατάξιμες μεταβλητές (Ordinal variables). Αυτές οι μεταβλητές μετριούνται σε κατηγορίες που μπορούν με κάποιον τρόπο να διαταχθούν. Για παράδειγμα, σε ένα ερωτηματολόγιο το επίπεδο συμφωνίας με μία δήλωση μπορεί να δηλώνεται ως «Καθόλου», «Λίγο», «Πολύ», «Απολύτως». Σε αυτήν την περίπτωση, το «Λίγο» ξέρουμε ότι είναι μικρότερο από το «Απολύτως», αν και δεν έχουμε εικόνα πόσο «απέχει» το «Λίγο» από το «Απολύτως». Στο Jamovi αυτές συμβολίζονται με μία σκάλα.
4. Συνεχείς μεταβλητές (Continuous variables). Αυτές οι μεταβλητές μετριούνται σε μία συνεχή αριθμητική κλίμακα. Για παράδειγμα, το ύψος και το βάρος των μαθητών σε μία

τάξη είναι συνεχείς μεταβλητές. Ένα ύψος 110 εκατοστών είναι μεγαλύτερο από ένα ύψος 100 εκατοστών. Συγχρόνως, ξέρουμε και πόσο μεγαλύτερο είναι το πρώτο ύψος από το δεύτερο (10 εκατοστά). Στο Jamovi αυτές συμβολίζονται με έαν χάρακα.

Αν δούμε τα στο συγκεκριμένο αρχείο, έχουμε μία μεταβλητή που ονομάζεται school\_id την οποία το Jamovi θεώρησε ονομαστική (Figure 4).

	school_id	child_id	child_ge...	child_age	child_rea...	child_kn...
1	38	0038004C	MALE	75.20	77.60	
2	38	0038005C	FEMALE	64.97	28.21	
3	38	0038010C	FEMALE	66.13	27.19	
4	38	0038011C	MALE	69.33	32.05	
5	38	0038012C	FEMALE	70.73	33.59	
6	38	0038015C	MALE	70.70	37.89	
7	38	0038017C	FEMALE	74.33	31.56	
8	38	0038018C	FEMALE	68.73	26.90	
9	38	0038019C	MALE	70.73	46.62	
10	38	0038020C	MALE	66.00	25.28	
11	38	0038021C	MALE	70.70	46.02	
12	78	0078001C	FEMALE	69.60	42.32	
13	78	0078003C	FEMALE	65.57	32.39	
14	78	0078004C	FEMALE	66.77	39.04	
15	78	0078005C	FEMALE	68.47	32.00	
16	78	0078006C	FEMALE	63.33	59.16	

Figure 4: Η μεταβλητή school\_id δεν θα έπρεπε να είναι ονομαστική καθώς μας δίνει έναν αναγωριστικό αριθμό για κάθε σχολείο.

Για να αλλάξουμε τον τύπο της μεταβλητής κάνουμε διπλό κλικ επάνω της και οδηγούμαστε στο μενού αλλαγής της μεταβλητής. Από εκεί διαλέγουμε την επιλογή ID και μετατρέπουμε τη school\_id από ονομαστική σε αναγνωριστική (Figure 5).

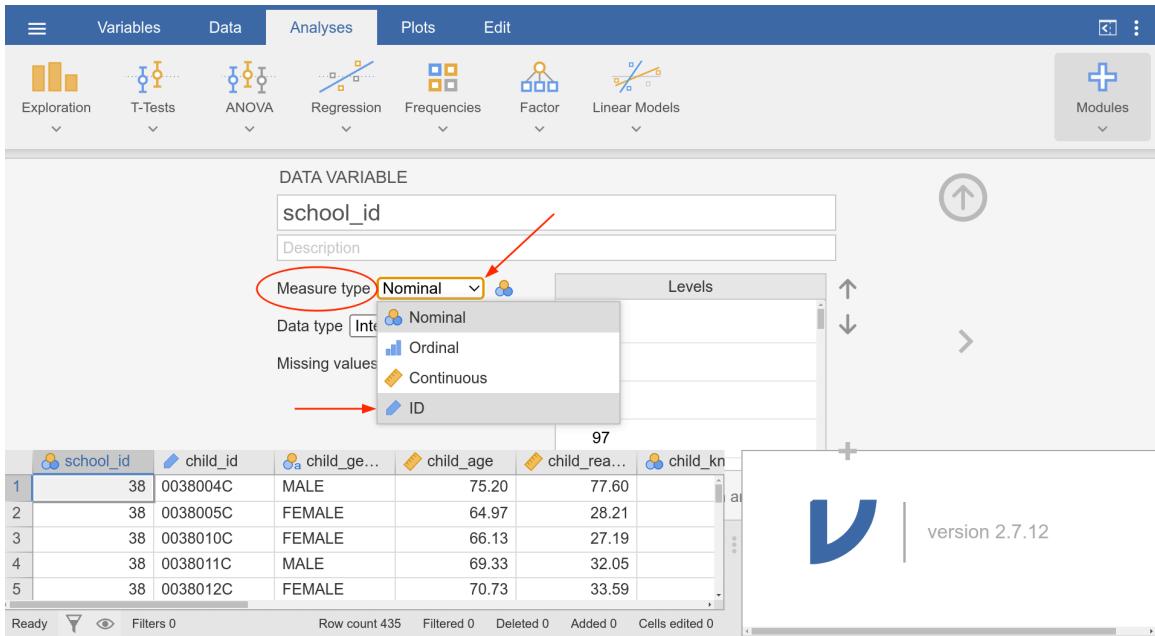


Figure 5: Αλλαγή τύπου της μεταβλητής school\_id σε αναγνωριστική.

Αν κοιτάξουμε και τις υπόλοιπες μεταβλητές, θα δούμε και άλλες περιπτώσεις που το Jamovi δεν έκανε καλή ερμηνεία του τύπου των μεταβλητών. Για παράδειγμα, η μεταβλητή parent\_read\_to\_child που μας περιγράφει τη συχνότητα μέσα στην εβδομάδα που οι γονείς διαβάζουν στο παιδί είναι ονομαστική ενώ κανονικά θα έπρεπε να είναι διατάξιμη. Οι τιμές της πηγαίνουν από το καθόλου έως κάθε μέρα.

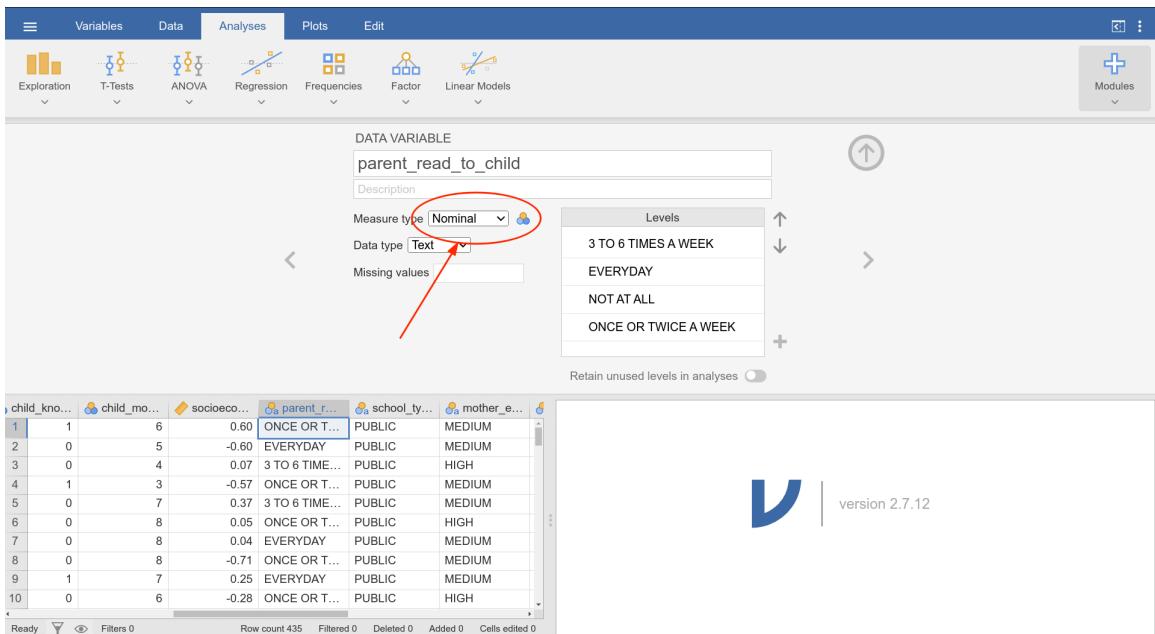
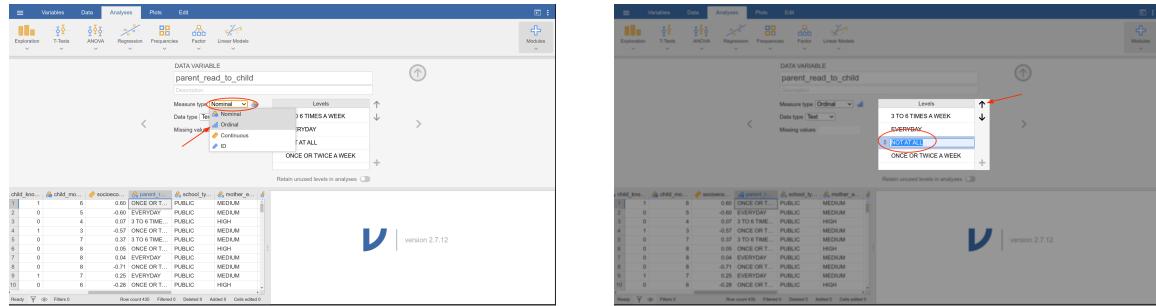


Figure 6: Η μεταβλητή parent\_read\_to\_child δεν θα έπρεπε να είναι ονομαστική αλλά διατάξιμη.

Η αλλαγή αυτή γίνεται σε δύο βήματα. Το πρώτο βήμα είναι το ίδιο με το Figure 5, και περιλαμβάνει την αλλαγή του τύπου της μεταβλητής. Στο δεύτερο βήμα πρέπει να αναδιατάξουμε τις τιμές της μεταβλητής, ώστε να αντικατοπτρίζουν τη διάταξη. Συνήθως, βάζομε τις αλλαγές σε αύξουσα σειρά, από τη μικρότερη τιμή στη μεγαλύτερη (Figure 7).



(a) Αλλαγή τύπου της μεταβλητής. (b) Αναδιάταξη σε αύξουσα σειρά.  
Figure 7: Διαδικασία φόρτωσης δεδομένων στο Jamovi.

Στο τέλος, η μεταβλητή `parent_read_to_child` θα πρέπει να έχει τη μορφή που φαίνεται στο Figure 8.

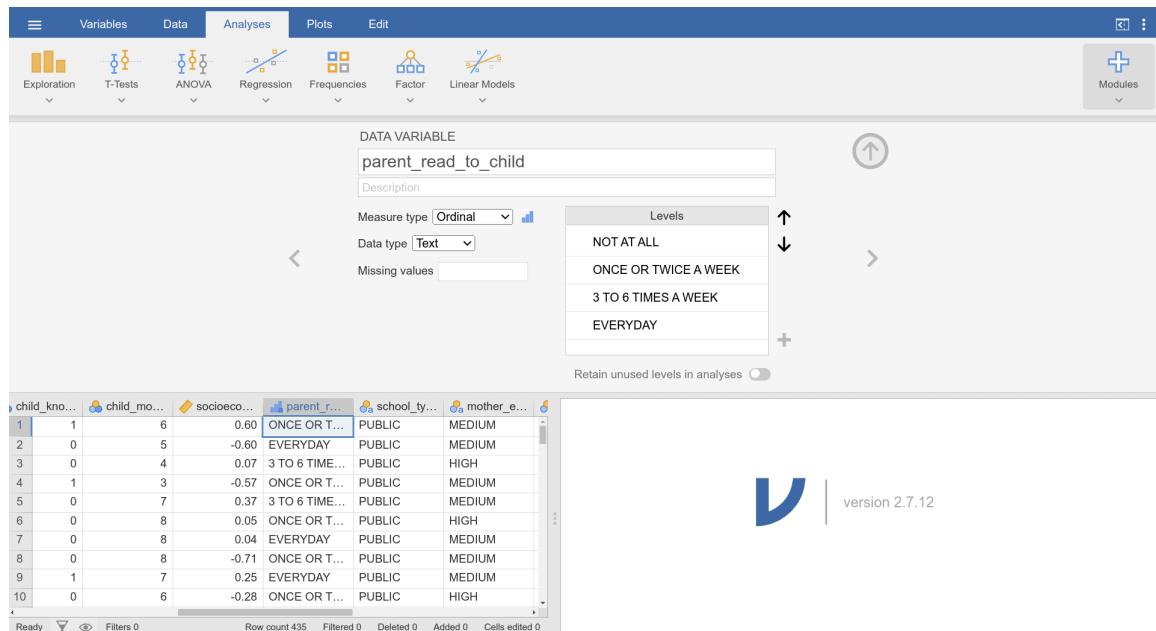


Figure 8: Η σωστή μορφή που παίρνει η μεταβλητή `parent_read_to_child` μετά τις αλλαγές.

Με τον ίδιο τρόπο αλλάζουμε και τον τύπο των μεταβλητών `mother_education` και `father_education`, οι οποίες πάλι θα έπρεπε να είναι διατάξιμες αλλά έχουν εισαχθεί ως ονομαστικές.