

Οδηγίες Χρήσης του Στατιστικού Πακέτου Jamovi

Μάθημα:

Ποσοτικές Μέθοδοι στην Εκπαιδευτική Έρευνα (0109Ε)

Αλέξανδρος Ρέκκας

Παιδαγωγικό Τμήμα Νηπιαγωγών, Πανεπιστήμιο Δυτικής Μακεδονίας

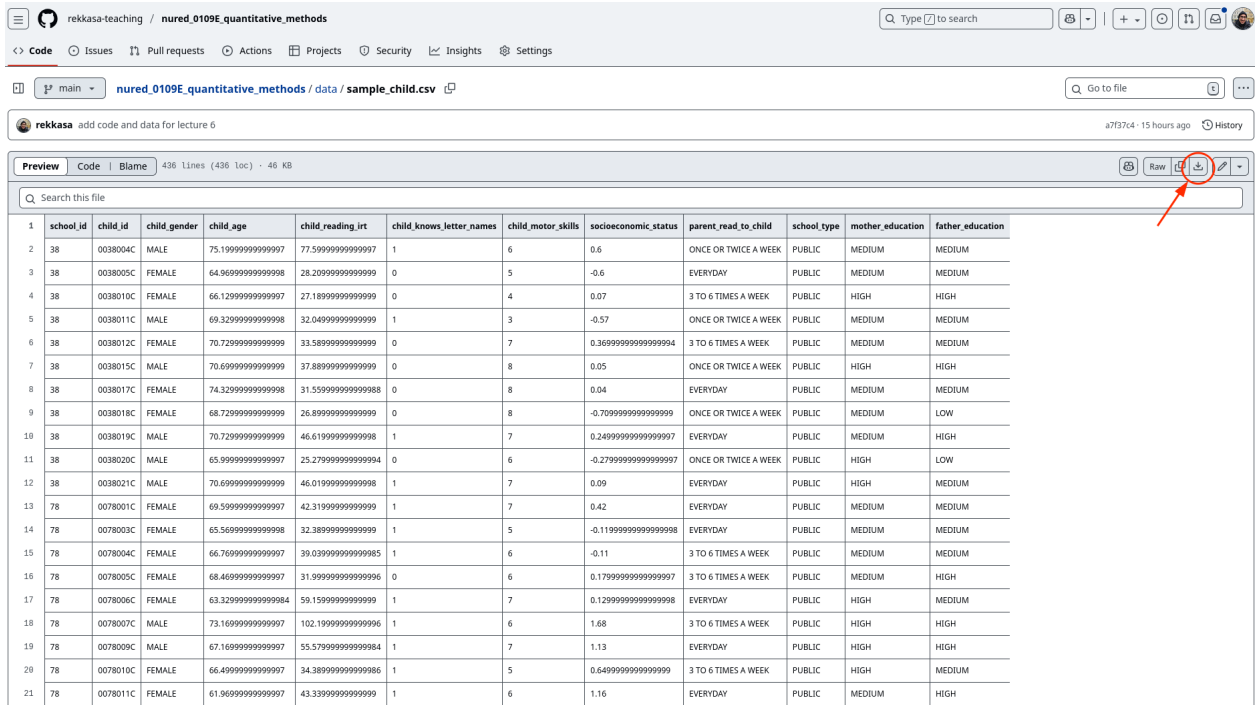
Περιεχόμενα

1	Εισαγωγή και επεξεργασία δεδομένων	5
1.1	Εισαγωγή δεδομένων	5
1.2	Αρχική επεξεργασία	6
2	Γραμμική παλινδρόμηση	11

1 Εισαγωγή και επεξεργασία δεδομένων

1.1 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.



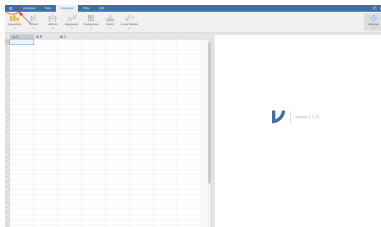
The screenshot shows a GitHub repository for 'rekksa-teaching / nured_0109E_quantitative_methods'. The file 'sample_child.csv' is selected, showing its content in a table format. The table has 11 columns: school_id, child_id, child_gender, child_age, child_reading_irt, child_knows_letter_names, child_motor_skills, socioeconomic_status, parent_read_to_child, school_type, mother_education, and father_education. The download icon is circled in red in the top right corner of the file viewer.

	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	MEDIUM
2	38	0038005C	FEMALE	64.90999999999998	28.20999999999999	0	5	-0.6	EVERYDAY	PUBLIC	MEDIUM	MEDIUM
3	38	0038010C	FEMALE	66.12999999999997	27.18999999999999	0	4	0.07	3 TO 6 TIMES A WEEK	PUBLIC	HIGH	HIGH
4	38	0038011C	MALE	69.32999999999998	32.04999999999999	1	3	-0.57	ONCE OR TWICE A WEEK	PUBLIC	MEDIUM	MEDIUM
5	38	0038012C	FEMALE	70.72999999999999	33.88999999999999	0	7	0.3699999999999994	3 TO 6 TIMES A WEEK	PUBLIC	MEDIUM	MEDIUM
6	38	0038015C	MALE	70.69999999999999	37.88999999999999	0	8	0.05	ONCE OR TWICE A WEEK	PUBLIC	HIGH	HIGH
7	38	0038017C	FEMALE	74.32999999999998	31.55999999999998	0	8	0.04	EVERYDAY	PUBLIC	MEDIUM	MEDIUM
8	38	0038018C	FEMALE	68.72999999999999	26.89999999999999	0	8	-0.7099999999999999	ONCE OR TWICE A WEEK	PUBLIC	MEDIUM	LOW
9	38	0038019C	MALE	70.72999999999999	46.61999999999998	1	7	0.2499999999999997	EVERYDAY	PUBLIC	MEDIUM	HIGH
10	38	0038020C	MALE	65.99999999999997	25.27999999999994	0	6	-0.2799999999999997	ONCE OR TWICE A WEEK	PUBLIC	HIGH	LOW
11	38	0038021C	MALE	70.69999999999999	46.01999999999998	1	7	0.09	EVERYDAY	PUBLIC	HIGH	MEDIUM
12	78	0078001C	FEMALE	69.59999999999997	42.31999999999999	1	7	0.42	EVERYDAY	PUBLIC	MEDIUM	MEDIUM
13	78	0078003C	FEMALE	65.56999999999998	32.38999999999999	1	5	-0.1199999999999998	EVERYDAY	PUBLIC	MEDIUM	MEDIUM
14	78	0078004C	FEMALE	66.76999999999997	39.039999999999985	1	6	-0.11	3 TO 6 TIMES A WEEK	PUBLIC	MEDIUM	MEDIUM
15	78	0078005C	FEMALE	68.40999999999997	31.999999999999996	0	6	0.1799999999999997	3 TO 6 TIMES A WEEK	PUBLIC	MEDIUM	HIGH
16	78	0078006C	FEMALE	63.329999999999984	59.15999999999999	1	7	0.1299999999999998	EVERYDAY	PUBLIC	HIGH	MEDIUM
17	78	0078007C	MALE	73.16999999999997	102.19999999999996	1	6	1.68	3 TO 6 TIMES A WEEK	PUBLIC	HIGH	HIGH
18	78	0078009C	MALE	67.16999999999997	55.579999999999984	1	7	1.13	EVERYDAY	PUBLIC	HIGH	HIGH
19	78	0078010C	FEMALE	66.49999999999997	34.389999999999986	1	5	0.6499999999999999	3 TO 6 TIMES A WEEK	PUBLIC	HIGH	MEDIUM
20	78	0078011C	FEMALE	61.96999999999997	43.33999999999999	1	6	1.16	EVERYDAY	PUBLIC	MEDIUM	HIGH

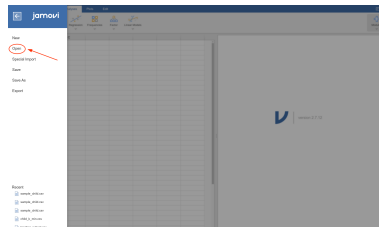
Σχήμα 1: Κουμπί download

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.

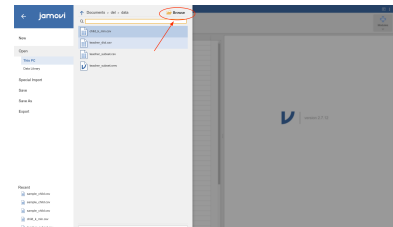
Αν έχετε ανοιχτό το πρόγραμμα Jamovi, μπορείτε να εισάγετε τα δεδομένα στο που κατεβάσατε ακολουθώντας τις εικόνες παρακάτω:



(a) Βήμα 1



(b) Βήμα 2



(c) Βήμα 3

Σχήμα 2: Διαδικασία φόρτωσης δεδομένων στο Jamovi.

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

	school_id	child_id	child_gender	child_age	child_readiness	child_knowledge	child_months	socioeconomic
1	38	0038004C	MALE	75.20	77.60	1	6	
2	38	0038005C	FEMALE	64.97	28.21	0	5	
3	38	0038010C	FEMALE	66.13	27.19	0	4	
4	38	0038011C	MALE	69.33	32.05	1	3	
5	38	0038012C	FEMALE	70.73	33.59	0	7	
6	38	0038015C	MALE	70.70	37.89	0	8	
7	38	0038017C	FEMALE	74.33	31.56	0	8	
8	38	0038018C	FEMALE	68.73	26.90	0	8	
9	38	0038019C	MALE	70.73	46.62	1	7	
10	38	0038020C	MALE	66.00	25.28	0	6	
11	38	0038021C	MALE	70.70	46.02	1	7	
12	78	0078001C	FEMALE	69.60	42.32	1	7	
13	78	0078003C	FEMALE	65.57	32.39	1	5	
14	78	0078004C	FEMALE	66.77	39.04	1	6	
15	78	0078005C	FEMALE	68.47	32.00	0	6	
16	78	0078006C	FEMALE	63.33	59.16	1	7	
17	78	0078007C	MALE	73.17	102.20	1	6	
18	78	0078009C	MALE	67.17	55.58	1	7	
19	78	0078010C	FEMALE	66.50	34.39	1	5	
20	78	0078011C	FEMALE	61.97	43.34	1	6	
21	78	0078013C	FEMALE	63.57	36.19	1	6	
22	78	0078014C	MALE	63.00	27.26	0	4	
23	78	0078015C	MALE	68.93	24.50	0	7	
24	78	0078016C	MALE	65.70	42.69	1	5	
25	78	0078018C	FEMALE	66.03	49.35	1	7	
26	78	0078020C	MALE	75.27	38.75	1	6	
27	78	0078021C	FEMALE	61.77	32.29	1	6	
28	78	0078022C	MALE	69.23	40.67	1	3	
29	78	0078023C	MALE	60.90	35.84	1	2	
30	80	0080002C	MALE	69.13	43.37	1	8	
31	80	0080003C	MALE	74.07	40.58	1	7	
32	80	0080005C	MALE	74.17	38.66	0	6	
33	80	0080007C	FEMALE	61.07	30.73	0	3	
34	80	0080008C	MALE	67.67	32.39	1	8	
35	80	0080009C	FEMALE	67.60	24.63	0	5	
36	80	0080010C	FEMALE	68.80	39.02	1	9	
37	80	0080011C	FEMALE	70.23	35.94	1	8	
38	80	0080013C	MALE	65.77	27.07	0	5	

Σχήμα 3: Τα δεδομένα έχουν φορτωθεί στο Jamoni με επιτυχία.

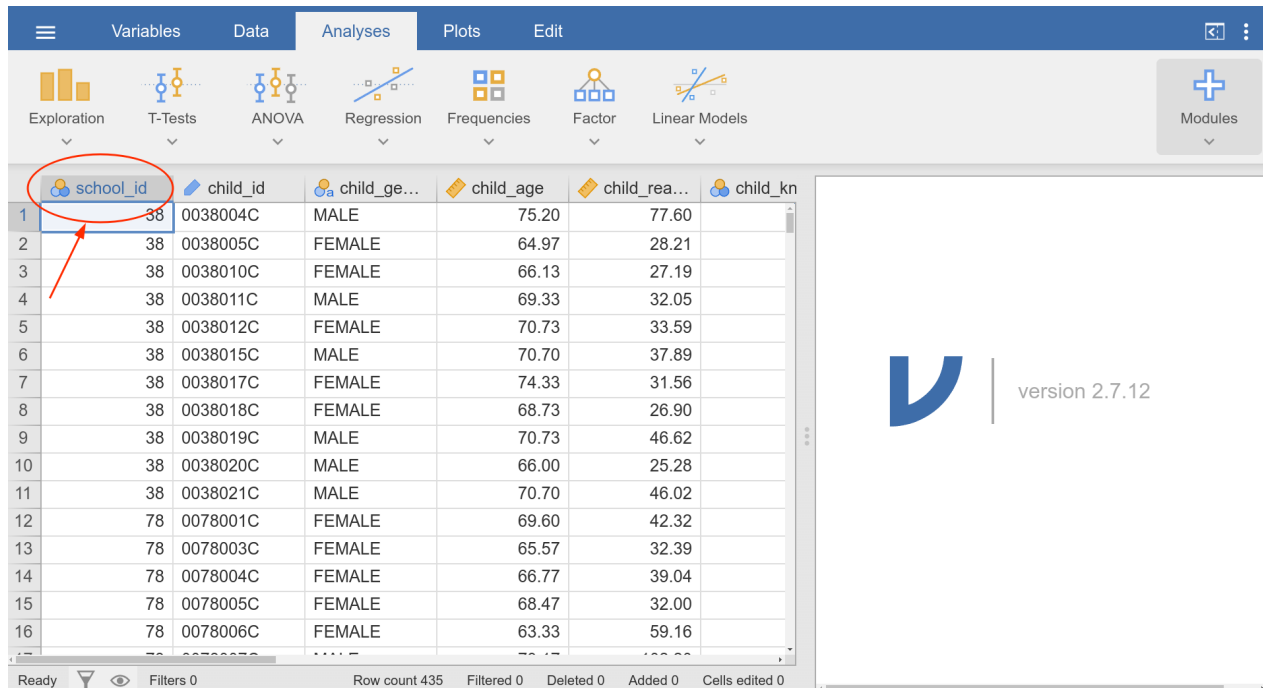
1.2 Αρχική επεξεργασία

Όταν εισάγουμε τα δεδομένα μας, το Jamoni 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 Jamoni 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 Jamoni αυτές συμβολίζονται με ένα μικρό ταμπελάκι.
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 Jamoni αυτές συμβολίζονται με τρεις τεμνόμενους κύκλους.
3. Διατάξιμες μεταβλητές (Ordinal variables). Αυτές οι μεταβλητές μετριοούνται 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 Jamoni αυτές συμβολίζονται με μία σκάλα.

4. Συνεχείς μεταβλητές (Continuous variables). Αυτές οι μεταβλητές μετριοούνται 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 Jamonι αυτές συμβολίζονται με έναν χάρακα.

Αν δούμε τα στο συγκεκριμένο αρχείο, έχουμε μία μεταβλητή που ονομάζεται `school_id` την οποία το Jamonι θεώρησε ονομαστική (Σχήμα 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	

Σχήμα 4: Η μεταβλητή `school_id` δεν θα έπρεπε να είναι ονομαστική καθώς μας δίνει έναν αναγνωριστικό 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 (Σχήμα 5).

DATA VARIABLE

school_id

Description

Measure type: **Nominal** (selected)

Data type: Integer

Missing values:

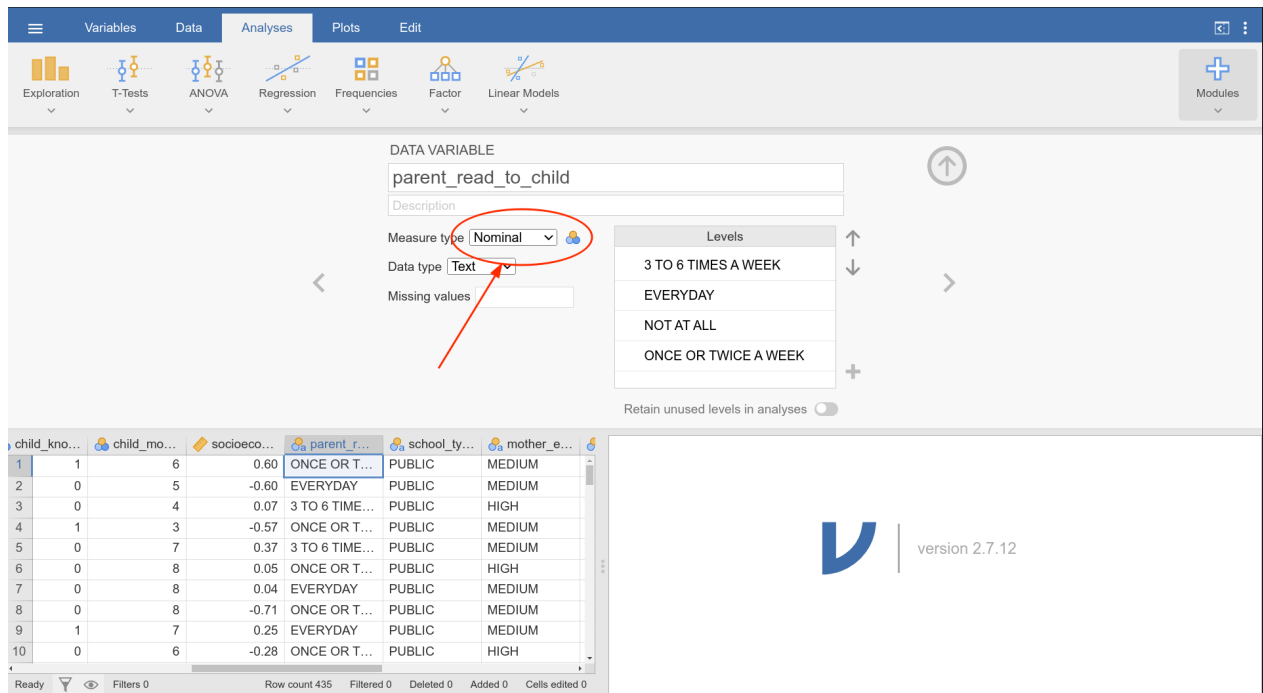
Levels:

school_id	child_id	child_gender	child_age	child_read_to_child	child_knowledge
38	0038004C	MALE	75.20	77.60	
38	0038005C	FEMALE	64.97	28.21	
38	0038010C	FEMALE	66.13	27.19	
38	0038011C	MALE	69.33	32.05	
38	0038012C	FEMALE	70.73	33.59	

version 2.7.12

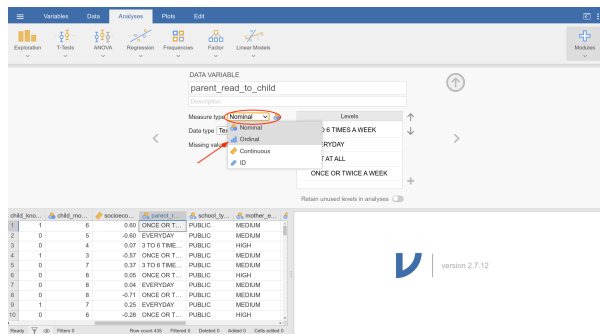
Σχήμα 5: Αλλαγή τύπου της μεταβλητής school_id σε αναγνωριστική.

Αν κοιτάξουμε και τις υπόλοιπες μεταβλητές, θα δούμε και άλλες περιπτώσεις που το Jamoni 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.

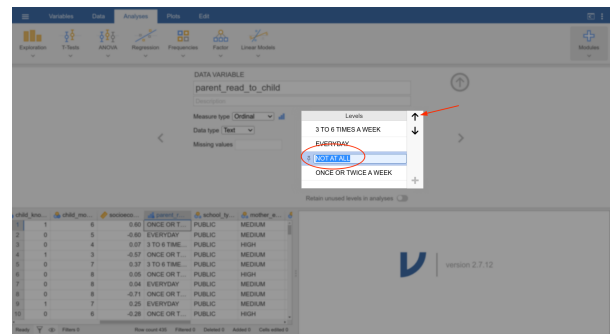


Σχήμα 6: Η μεταβλητή parent_read_to_child δεν θα έπρεπε να είναι ονομαστική αλλά διατάξιμη.

Η αλλαγή αυτή γίνεται σε δύο βήματα. Το πρώτο βήμα είναι το ίδιο με το Σχήμα 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, βάζουμε τις αλλαγές σε αύξουσα σειρά, από τη μικρότερη τιμή στη μεγαλύτερη (Σχήμα 7).



(a) Αλλαγή τύπου της μεταβλητής.



(b) Αναδιάταξη σε αύξουσα σειρά.

Σχήμα 7: Διαδικασία φόρτωσης δεδομένων στο Jamovi.

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

DATA VARIABLE

parent_read_to_child

Description

Measure type **Ordinal**

Data type **Text**

Missing values

Levels

- NOT AT ALL
- ONCE OR TWICE A WEEK
- 3 TO 6 TIMES A WEEK
- EVERYDAY

Retain unused levels in analyses

child_kno...	child_mo...	socioeco...	parent_r...	school_ty...	mother_e...	
1	1	6	0.60	ONCE OR T...	PUBLIC	MEDIUM
2	0	5	-0.60	EVERYDAY	PUBLIC	MEDIUM
3	0	4	0.07	3 TO 6 TIME...	PUBLIC	HIGH
4	1	3	-0.57	ONCE OR T...	PUBLIC	MEDIUM
5	0	7	0.37	3 TO 6 TIME...	PUBLIC	MEDIUM
6	0	8	0.05	ONCE OR T...	PUBLIC	HIGH
7	0	8	0.04	EVERYDAY	PUBLIC	MEDIUM
8	0	8	-0.71	ONCE OR T...	PUBLIC	MEDIUM
9	1	7	0.25	EVERYDAY	PUBLIC	MEDIUM
10	0	6	-0.28	ONCE OR T...	PUBLIC	HIGH

version 2.7.12

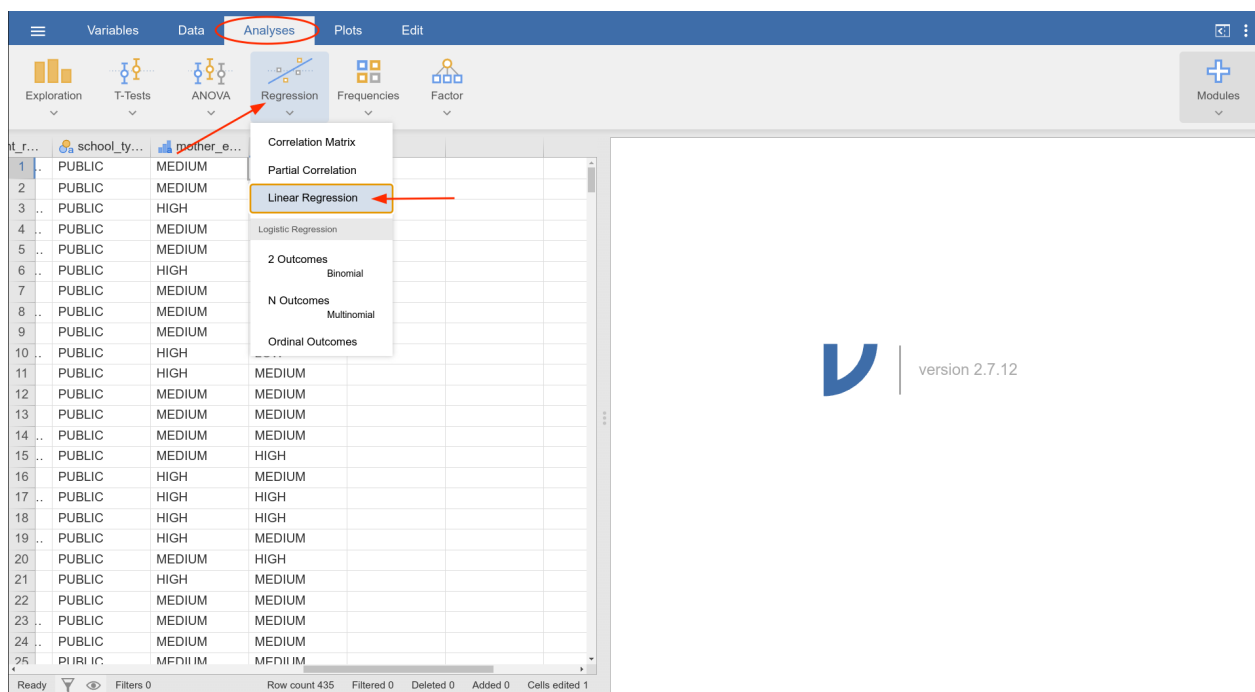
Σχήμα 8: Η σωστή μορφή που παίρνει η μεταβλητή parent_read_to_child μετά τις αλλαγές. Με τον ίδιο τρόπο αλλάζουμε και τον τύπο των μεταβλητών mother_education και father_education, οι οποίες πάλι θα έπρεπε να είναι διατάξιμες αλλά έχουν εισαχθεί ως ονομαστικές.

2 Γραμμική παλινδρόμηση

Έστω τώρα ότι θέλουμε να διαμορφώσουμε ένα μοντέλο γραμμικής παλινδρόμησης για τη συνεχή μεταβλητή `child_reading_irt` που μετρά την αναγνωστική ικανότητα των παιδιών στο νηπιαγωγείο. Για την πρόβλεψη της τιμής της `child_reading_irt` θα χρησιμοποιήσουμε τις εξής μεταβλητές του δείγματος:

- `child_gender`: το φύλο του παιδιού.
- `child_age`: η ηλικία του παιδιού μετρημένη σε μήνες.
- `parent_read_to_child`: η συχνότητα μέσα στην εβδομάδα που οι γονείς διαβάζουν στο παιδί.
- `socioeconomic_status`: η κοινωνικο/οικονομική κατάσταση της οικογένειας.
- `school_type`: το είδος του σχολείου που πηγαίνει το παιδί (ιδιωτικό / δημόσιο).
- `mother_education`: το μορφωτικό επίπεδο της μητέρας.
- `father_education`: το μορφωτικό επίπεδο του πατέρα.

Για να τρέξουμε την παραπάνω γραμμική παλινδρόμηση (και αφού έχουμε εφαρμόσει την αρχική επεξεργασία) από την καρτέλα των Αναλύσεων (**Analyses**), διαλέγουμε την παλινδρόμηση (**Regression**) και στο μενού που ανοίγει διαλέγουμε την γραμμική παλινδρόμηση (**Linear regression**), όπως φαίνεται και στο Σχήμα 9. Εν συντομία στο Jamoni, ακολουθούμε **Analyses > Regression > Linear Regression**.



Σχήμα 9: Η σωστή μορφή που παίρνει η μεταβλητή `parent_read_to_child` μετά τις αλλαγές. Το Jamoni διαχωρίζει τις μεταβλητές που χρησιμοποιούνται σε μία γραμμική παλινδρόμηση σε τρεις κατηγορίες:

- Την εξαρτημένη μεταβλητή, δηλαδή τη μεταβλητή που θέλουμε να προβλέψουμε με το μοντέλο, την οποία το πακέτο ονομάζει *Dependent Variable*.

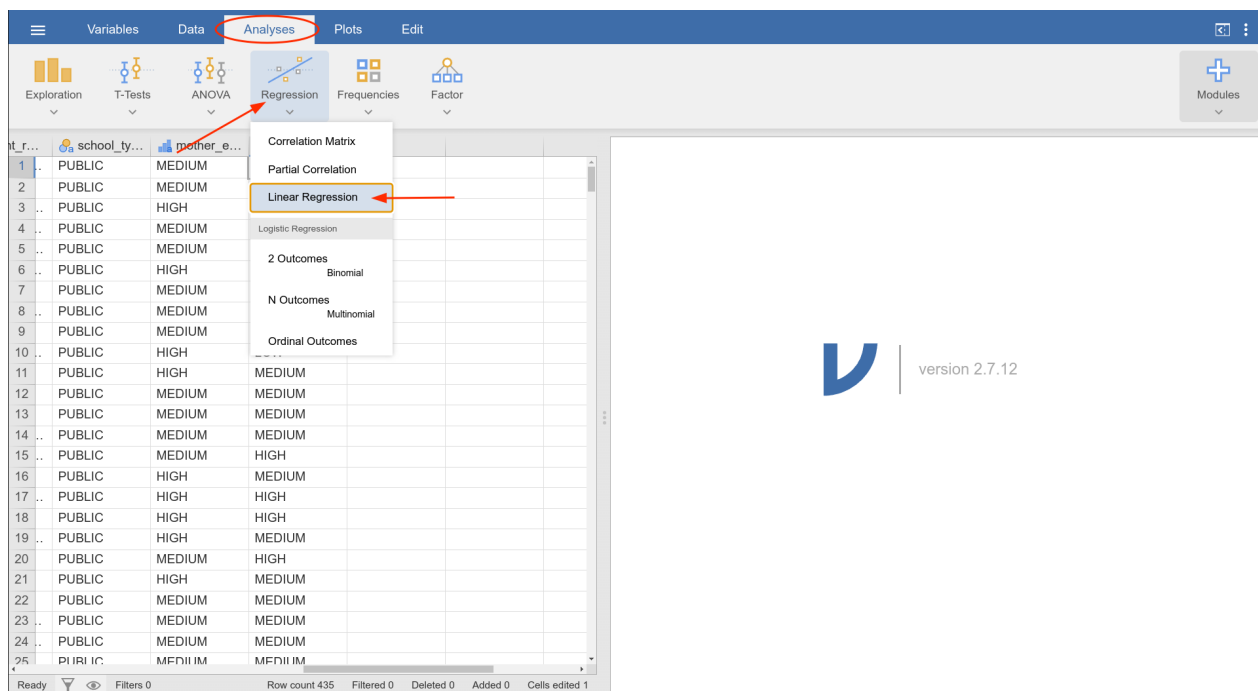
- Τις συνεχείς μεταβλητές, δηλαδή τις μεταβλητές των οποίων οι τιμές είναι αριθμοί, τις οποίες το πακέτο ονομάζει *Covariates*.
- Τις κατηγορικές μεταβλητές, δηλαδή τις μεταβλητές των οποίων οι τιμές είναι κατηγορίες (ονομαστικές, διατάξιμες), τις οποίες το πακέτο ονομάζει *Factors*.

Για να ορίσουμε το μοντέλο της γραμμικής παλινδρόμησης πρέπει σύρουμε και να αφήσουμε τις μεταβλητές που θα χρησιμοποιήσουμε στα σωστά σημεία (Σχήμα 10):

- Η μεταβλητή `child_reading_irt` είναι η εξαρτημένη μεταβλητή (dependent variable), η μεταβλητή δηλαδή που θέλουμε να προβλέψουμε και άρα θα πάει στο κουτί *Dependent Variable*.

Προσοχή: Μόνο μία εξαρτημένη μεταβλητή μπορεί να υπάρχει στη γραμμική παλινδρόμηση.

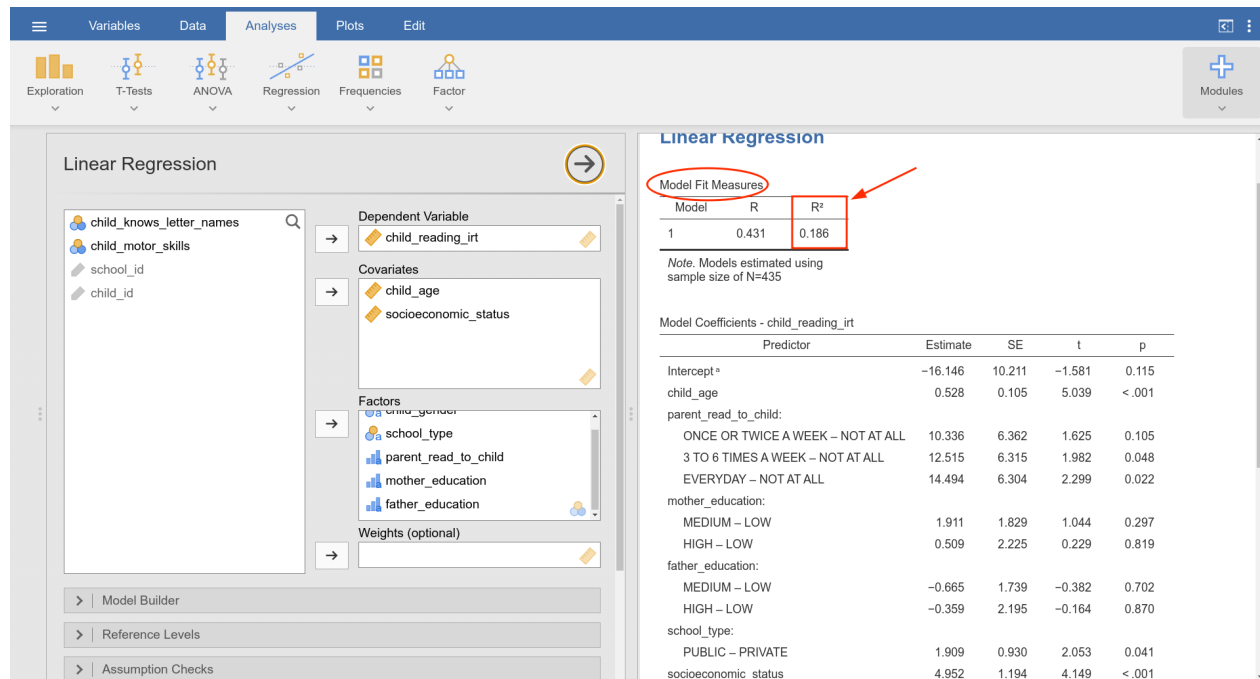
- Οι μεταβλητές `child_age` και `socioeconomic_status` είναι συνεχείς (αριθμητικές), άρα θα πάνε στο κουτί *Covariates*.
- Οι μεταβλητές `child_gender`, `parent_read_to_child`, `school_type`, `mother_education` και `father_education` είναι όλες κατηγορικές (ορίζουν κατηγορίες) και άρα θα πάνε στο κουτί *Factors*.



Σχήμα 10: Ορισμός του μοντέλου γραμμικής παλινδρόμησης.

Στο δεξί παράθυρο του Jamoni εμφανίζονται τα αποτελέσματα της γραμμικής παλινδρόμησης. Αν κάνουμε αλλαγές στον ορισμό του μοντέλου, τα αποτελέσματα ενημερώνονται αυτόματα. Στον πρώτο πίνακα των αποτελεσμάτων έχουμε τα μέτρα επίδοσης του μοντέλου παλινδρόμησης που διαλέξαμε (Σχήμα 11). Βλέπουμε ότι η προεπιλογή είναι να παρουσιάζονται μόνο το R^2 και η τετραγωνική του ρίζα. Στη συγκεκριμένη περίπτωση, το R^2 είναι αρκετά χαμηλό (0.186). Δηλαδή, μόνο περίπου το 18.6% της συνολικής διασποράς

της εξαρτημένης μεταβλητής (*child_reading_irt*) εξηγείται από το μοντέλο της γραμμικής παλινδρόμησης που υπολογίσαμε.

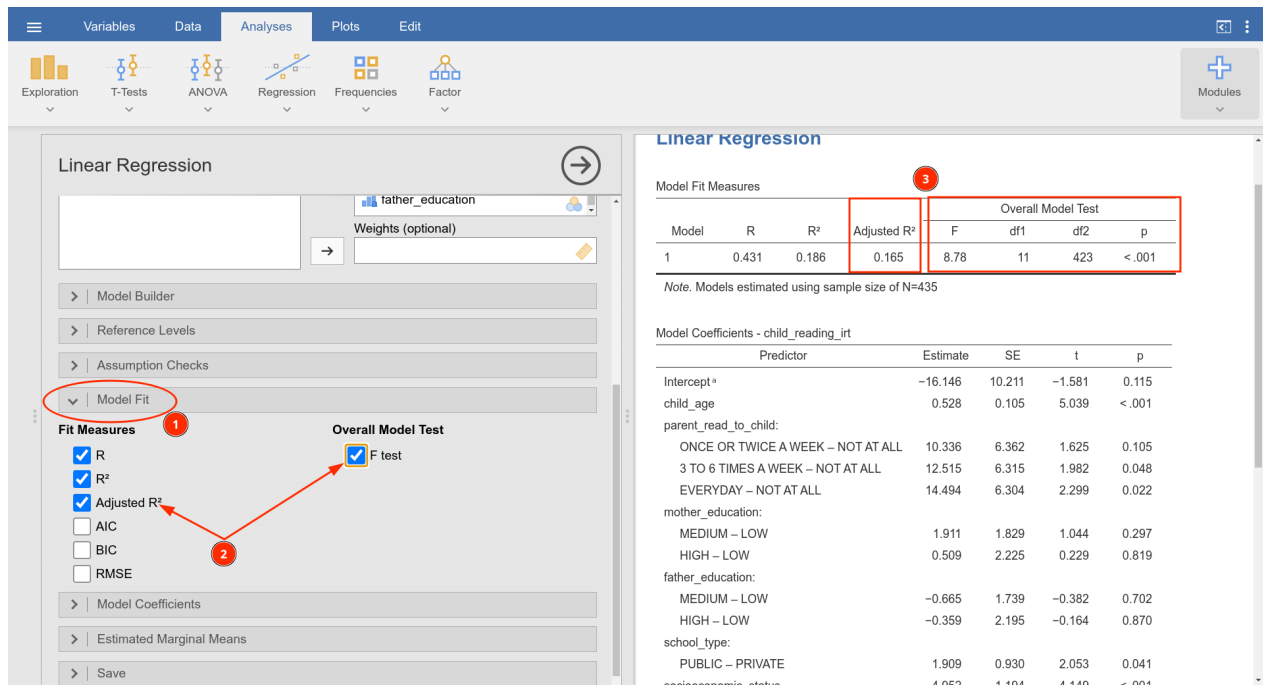


Σχήμα 11: Ορισμός του μοντέλου γραμμικής παλινδρόμησης.

Αν θέλουμε να δούμε και τα αποτελέσματα και άλλων μετρικών που εξετάζουν την ποιότητα του μοντέλου που υπολογίσαμε μπορούμε να τα προσθέσουμε πατώντας στο μενού **Model Fit**. Από εκεί έχει μία σειρά από επιλογές που μπορούμε να διαλέξουμε. Για παράδειγμα, μπορούμε να υπολογίσουμε το προσαρμοσμένο R^2 διαλέγοντας και την επιλογή **Adjusted R^2** (Σχήμα 12). Το προσαρμοσμένο R^2 προσθέτει ένα κόστος στη συμπερίληψη μεταβλητών στο μοντέλο με στόχο να διασφαλιστεί ότι οι επιπλέον μεταβλητές εξηγούν σημαντικό μέρος της ανεξήγητης διασποράς του μοντέλου χωρίς αυτές.

Είδαμε ότι το R^2 του μοντέλου είναι αρκετά χαμηλό. Άρα έχουμε λόγο να ανησυχούμε αν όντως έχει νόημα να κάνουμε όλη αυτή την άσκηση υπολογισμού το μοντέλου γραμμικής παλινδρόμησης. Με άλλα λόγια, αν το μοντέλο το οποίο υπολογίσαμε μας δίνει καλύτερες προβλέψεις για την αναγνωστική ικανότητα των μαθητών (*child_reading_irt*) από ό,τι αν χρησιμοποιούσαμε μόνο την μέση τιμή της χωρίς καμία άλλη πληροφορία. Έχουμε δει ότι τον έλεγχο αυτό τον πραγματοποιούμε με τη χρήση του συνολικού F -test. Αυτός ο έλεγχος μπορεί να συμπεριληφθεί στα αποτελέσματα χρησιμοποιώντας το ίδιο μενού (**Model Fit**) (Σχήμα 12).

Στη συγκεκριμένη περίπτωση, βλέπουμε ότι το συνολικό F -test έχει τιμή $F = 8.78$ με $p < 0.001$. Άρα μπορούμε να απορρίψουμε την H_0 ότι όλα τα β του μοντέλου είναι 0 και να δεχτούμε ότι έστω και ένα από τα β είναι διάφορο του μηδενός. Άρα, παρά τη χαμηλή επίδοση του μοντέλου παλινδρόμησης ως προς το R^2 , αυτό έχει μία αξία, δηλαδή μας επιτρέπει να κάνουμε καλύτερες προβλέψεις από το να χρησιμοποιούσαμε μόνο τον μέσο όρο της *child_reading_irt*.



Σχήμα 12: Πρόσθεση του προσαρμοσμένου R^2 και συνολικού F -test στα αποτελέσματα της αξιολόγησης του μοντέλου.

Το βασικό κομμάτι, όμως, των αποτελεσμάτων της γραμμικής παλινδρόμησης παρουσιάζεται στον παρακάτω πίνακα των αποτελεσμάτων και έχει τίτλο *Model Coefficients*. Το μοντέλο που θέλαμε να υπολογίσουμε είχε τη μορφή:

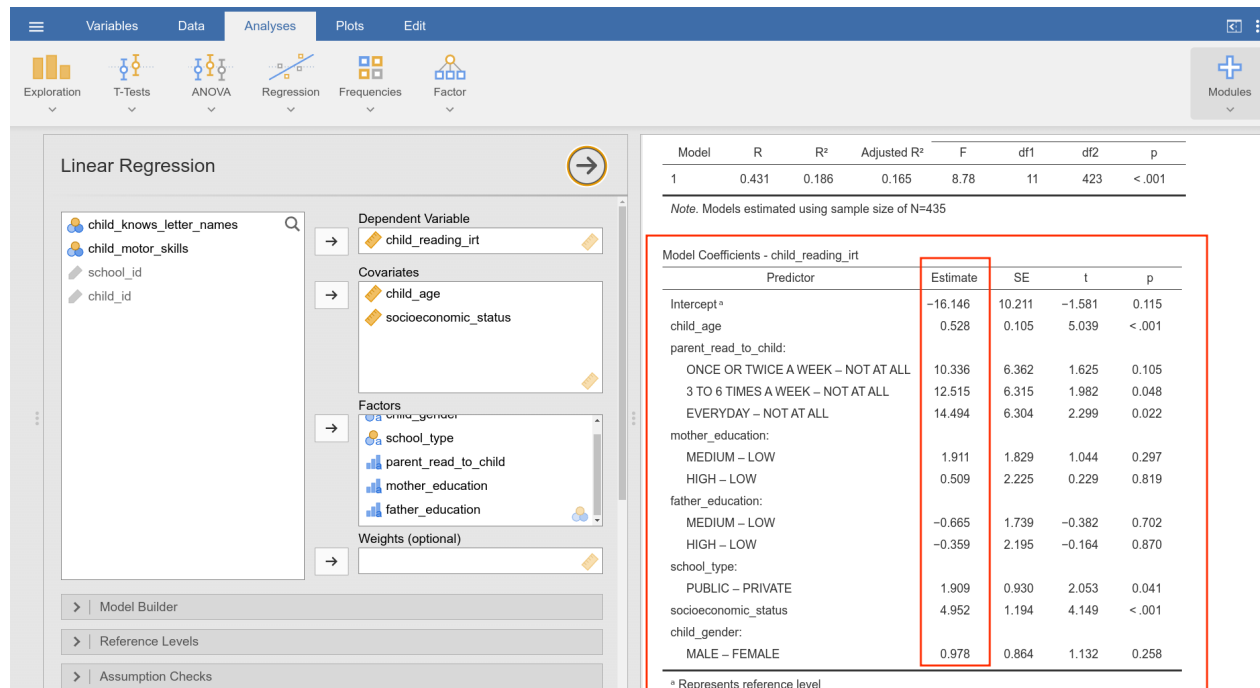
$$\begin{aligned} \text{child_reading_rt} = & \beta_0 + \beta_1 \times \text{child_age} + \beta_2 \times \text{socioeconomic_status} \\ & + \beta_3 \times \text{child_gender} + \beta_4 \times \text{mother_education} \\ & + \beta_5 \times \text{father_education} + \beta_6 \times \text{school_type} \\ & + \beta_7 \times \text{parent_read_to_child} \end{aligned}$$

Στο Σχήμα 13, όμως, βλέπουμε πολύ περισσότερα αποτελέσματα από τα 8 που φαίνονται στην παραπάνω εξίσωση. Αυτό συμβαίνει γιατί το Jamonι χωρίζει τις κατηγορικές μεταβλητές σε βοηθητικές μεταβλητές. Για τον χωρισμό σε βοηθητικές μεταβλητές χρησιμοποιείται ένα επίπεδο της κατηγορικής μεταβλητής ως σημείο αναφοράς. Στο Jamonι αυτό είναι το πρώτο επίπεδο. Για παράδειγμα, στην περίπτωση του φύλου δημιουργείται η βοηθητική μεταβλητή $X_{\text{child_gender} = \text{MALE}}$, η οποία παίρνει την τιμή 1 όταν το φύλο είναι αγόρι (MALE) και 0 στην αντίθετη περίπτωση.

Ομοίως, στην περίπτωση του μορφωτικού επιπέδου της μητέρας (LOW, MEDIUM, HIGH) δημιουργούνται 2 βοηθητικές μεταβλητές:

- $X_{\text{mother_education} = \text{MEDIUM}}$, η οποία παίρνει την τιμή 1 όταν το μορφωτικό επίπεδο της μητέρας είναι MEDIUM και 0 σε αντίθετη περίπτωση.
- $X_{\text{mother_education} = \text{HIGH}}$, η οποία παίρνει την τιμή 1 όταν το μορφωτικό επίπεδο της μητέρας είναι HIGH και 0 σε αντίθετη περίπτωση.

Τι γίνεται όταν το μορφωτικό επίπεδο της μητέρας είναι χαμηλό (LOW), δηλαδή το επίπεδο αναφοράς; Τότε και οι δύο βοηθητικές μεταβλητές είναι 0 και η επίδραση μορφωτικού επιπέδου της μητέρας απορροφάται από το β_0 . Γενικά, όταν έχουμε να κάνουμε με κατηγορικές μεταβλητές, το Jamoní θα δημιουργεί για κάθε μία από αυτές ένα πλήθος βοηθητικών μεταβλητών που παίρνουν τις τιμές 0 και 1. Το πλήθος αυτό θα είναι όσο είναι το πλήθος των επιπέδων της κατηγορικής μεταβλητής μείον 1.



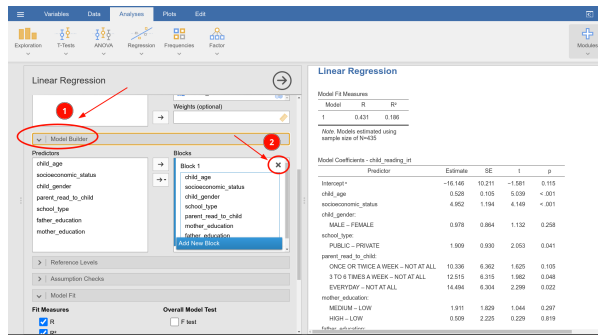
Σχήμα 13: Οι συντελεστές του μοντέλου της γραμμικής παλινδρόμησης που εκτιμήθηκαν χρησιμοποιώντας το δείγμα των 435 μαθητών των δεδομένων μας.

Στο αρχικό μοντέλο συμπεριλάβαμε, μεταξύ άλλων, τις μεταβλητές που περιγράφουν την κοινωνικο/οικονομική κατάσταση της οικογένειας (socioeconomic_status) και μεταβλητές για το μορφωτικό επίπεδο των γονέων καθώς και το είδος του σχολείου που πηγαίνουν τα παιδιά (mother_education, father_education, school_type). Έχει νόημα να συμπεριλάβουμε τις τελευταίες τρεις μεταβλητές ή κάνουμε πάνων-κάτω την ίδια δουλειά απλώς βάζοντας στο μοντέλο την κοινωνικο/οικονομική κατάσταση; Για να το ελέγξουμε αυτό πρέπει να πραγματοποιήσουμε το μερικό F -test. Στην πράξη, το μερικό F -test συγκρίνει δύο μοντέλα, ένα με όλες τις μεταβλητές με αυτό που δεν συμπεριλαμβάνει τις τρεις μεταβλητές, δηλαδή κάνοντας την υπόθεση ότι οι συντελεστές αυτών των τριών μεταβλητών (τα β) είναι 0.

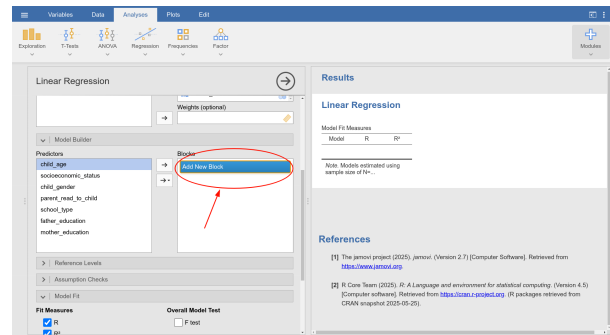
Στο Jamoní, για να πραγματοποιήσουμε το συγκεκριμένο έλεγχο πρέπει να διαλέξουμε το μενού Model Builder και να ορίσουμε εκεί τα δύο μοντέλα που είπαμε προηγουμένως, βάζοντας το καθένα σε ένα ξεχωριστό block. Στο block 1 βάζουμε τις μεταβλητές του απλούστερου μοντέλου, δηλαδή την ηλικία, την κοινωνικο/οικονομική κατάσταση, το φύλο του παιδιού και τη συχνότητα ανάγνωσης από τους γονείς (child_age, socioeconomic_status, child_gender, parent_read_to_child). Στο block 2 βάζουμε τις επιπλέον μεταβλητές που θέλουμε να ελέγξουμε αν έχει νόημα να συμπεριληφθούν στον μοντέλο, δηλαδή το μορφω-

τικό επίπεδο των γονέων και το είδος του σχολείου (mother_education, father_education, school_type).

Συνήθως, είναι πιο εύκολο πρώτα να καθαρίσουμε τα Block που υπάρχουν ήδη συμπληρωμένα. Όταν το κάνετε αυτό, τα αποτελέσματα στη δεξιά μεριά της οθόνης θα χαθούν (Σχήμα 14). Στο επόμενο βήμα θα συμπληρωθούν ξανά με τη δημιουργία των δύο μοντέλων παλινδρόμησης που έχουμε περιγράψει παραπάνω.



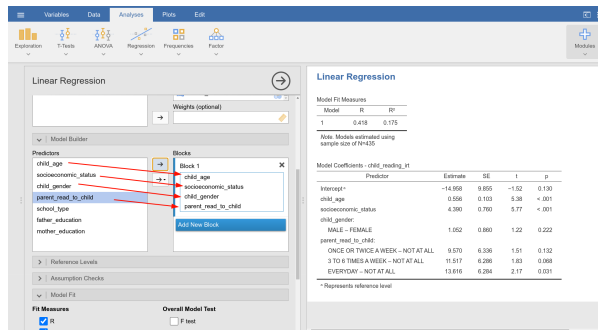
(a) Διαγραφή Block 1.



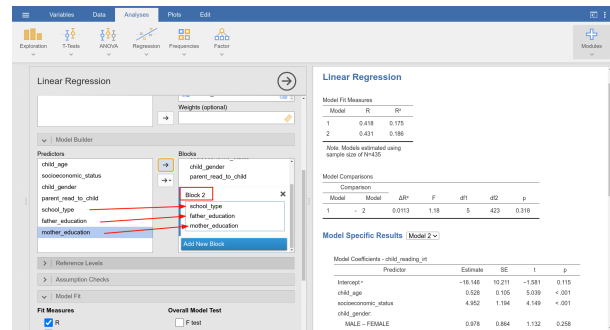
(b) Δημιουργία κενού Block 1

Σχήμα 14: Διαδικασία καθαρισμού υφιστάμενων Block.

Στη συνέχεια, δημιουργούμε ξανά τα Block όπως φαίνεται στο Σχήμα 15.



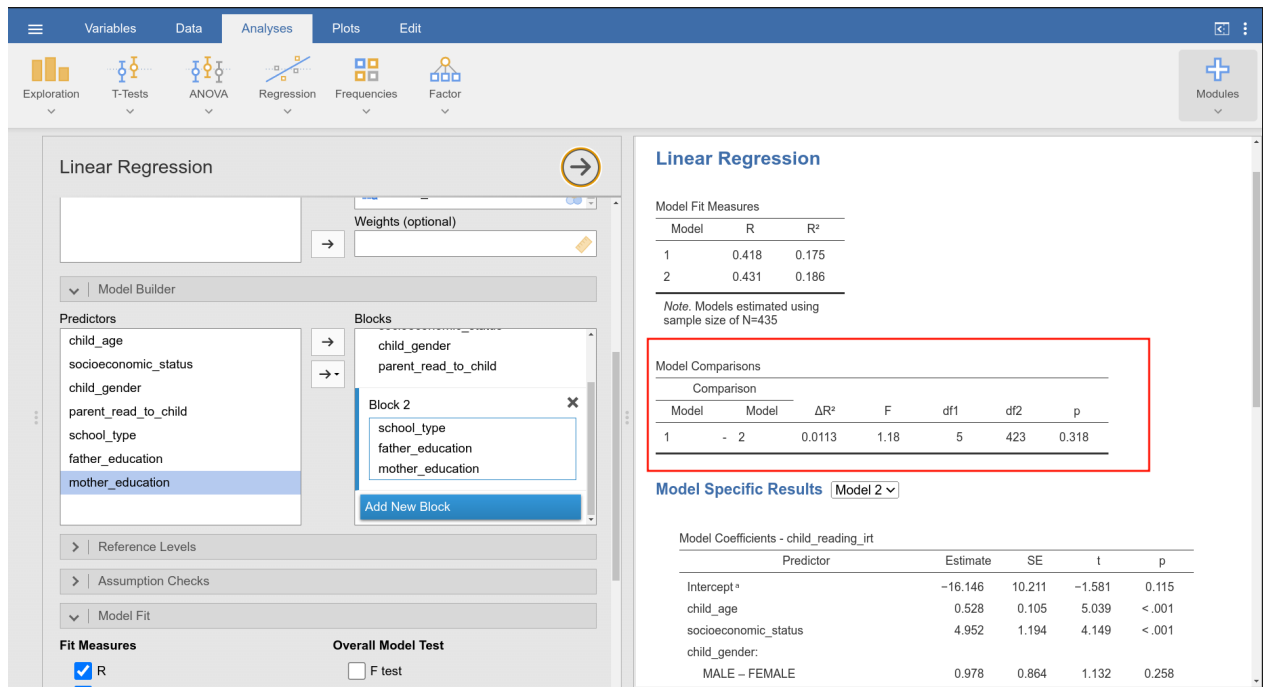
(a) Δημιουργία Block 1.



(b) Δημιουργία Block 2

Σχήμα 15: Διαδικασία δημιουργίας των νέων Block για τη διενέργεια του μερικού F -test.

Τα αποτελέσματα του μερικού F -test εμφανίζονται στη δεξιά μεριά της οθόνης (Σχήμα 16). Στη συγκεκριμένη περίπτωση, ελέγχουμε την υπόθεση (H_0) ότι οι συντελεστές των mother_education, father_education και school_type είναι 0. Στα αποτελέσματα βλέπουμε ότι η τιμή της F είναι 1.18 και το p είναι 0.318 (πολύ μεγαλύτερο από το 0.05 που χρησιμοποιούμε συνήθως για να συμπεράνουμε στατιστική σημαντικότητα). Άρα δεν μπορούμε να απορρίψουμε την H_0 και επομένως δεχόμαστε ότι οι συντελεστές των τριών αυτών μεταβλητών είναι 0. Άρα ένα απλούστερο μοντέλο, χωρίς τις τρεις αυτές μεταβλητές κάνει σχεδόν το ίδιο καλή δουλειά με το πλήρες μοντέλο.



Σχήμα 16: Τα αποτελέσματα του μερικού F -test.

