

# **Οδηγίες Χρήσης του Στατιστικού Πακέτου Jamovi**

**Μάθημα:**

**Ποσοτικές Μέθοδοι στην Εκπαιδευτική Έρευνα (0109Ε)**

Αλέξανδρος Ρέκκας

Παιδαγωγικό Τμήμα Νηπιαγωγών, Πανεπιστήμιο Δυτικής Μακεδονίας



# **Περιεχόμενα**

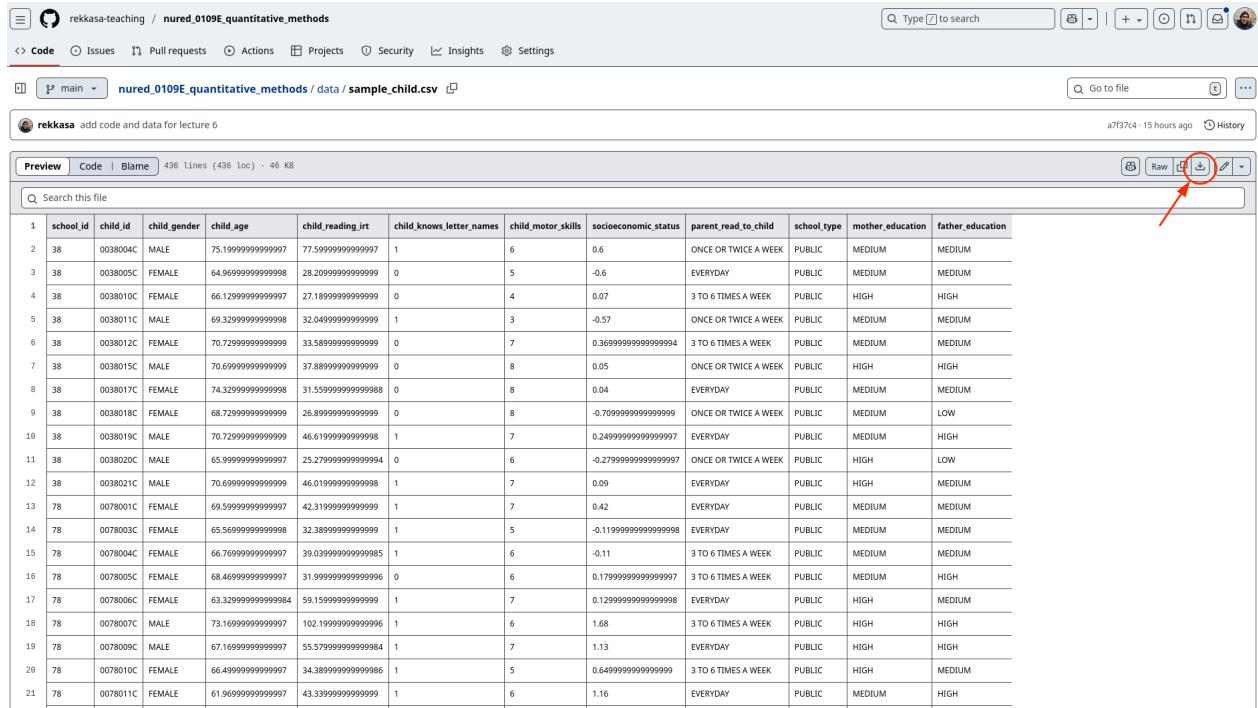
1 Εισαγωγή και επεξεργασία δεδομένων .....	5
1.1 Εισαγωγή δεδομένων .....	5
1.2 Αρχική επεξεργασία .....	6
2 Γραμμική παλινδρόμηση .....	10



# 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.

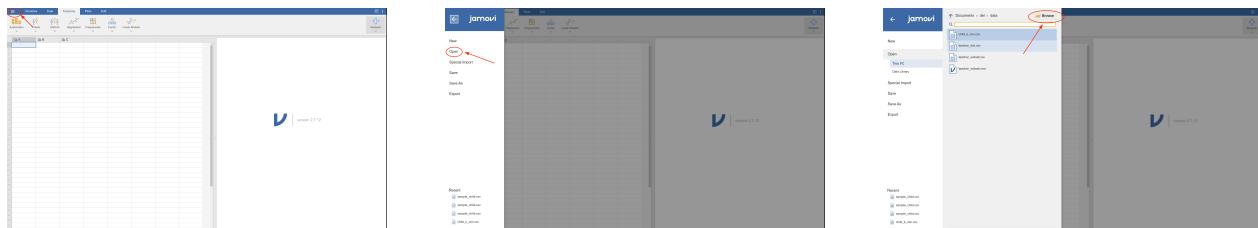


1	school_id	child_id	child_gender	child_age	child_reading_irt	child_knows_letter_names	child_motor_skills	socioeconomic_status	parent_read_to_child	school_type	mother_education	father_education
2	38	0038004C	MALE	75.19999999999997	77.59999999999997	1	6	0.6	ONCE OR TWICE A WEEK	PUBLIC	MEDIUM	MEDIUM
3	38	0038005C	FEMALE	64.69999999999998	28.20999999999999	0	5	-0.6	EVERYDAY	PUBLIC	MEDIUM	MEDIUM
4	38	0038010C	FEMALE	66.12999999999997	27.18999999999999	0	4	0.07	3 TO 6 TIMES A WEEK	PUBLIC	HIGH	HIGH
5	38	0038011C	MALE	69.32999999999998	32.04999999999999	1	3	-0.57	ONCE OR TWICE A WEEK	PUBLIC	MEDIUM	MEDIUM
6	38	0038012C	FEMALE	70.72999999999999	33.58999999999999	0	7	0.3699999999999994	3 TO 6 TIMES A WEEK	PUBLIC	MEDIUM	MEDIUM
7	38	0038015C	MALE	70.69999999999999	37.88999999999999	0	8	0.05	ONCE OR TWICE A WEEK	PUBLIC	HIGH	HIGH
8	38	0038017C	FEMALE	74.32999999999998	31.55999999999998	0	8	0.04	EVERYDAY	PUBLIC	MEDIUM	MEDIUM
9	38	0038018C	FEMALE	68.72999999999999	26.89999999999999	0	8	-0.7099999999999997	ONCE OR TWICE A WEEK	PUBLIC	MEDIUM	LOW
10	38	0038019C	MALE	70.72999999999999	46.61999999999998	1	7	0.2499999999999997	EVERYDAY	PUBLIC	MEDIUM	HIGH
11	38	0038020C	MALE	65.99999999999997	25.27999999999997	0	6	-0.2799999999999997	ONCE OR TWICE A WEEK	PUBLIC	HIGH	LOW
12	38	0038021C	MALE	70.69999999999999	46.01999999999998	1	7	0.09	EVERYDAY	PUBLIC	HIGH	MEDIUM
13	78	0078001C	FEMALE	69.59999999999997	42.31999999999999	1	7	0.42	EVERYDAY	PUBLIC	MEDIUM	MEDIUM
14	78	0078003C	FEMALE	65.56999999999998	32.38999999999999	1	5	-0.1199999999999998	EVERYDAY	PUBLIC	MEDIUM	MEDIUM
15	78	0078004C	FEMALE	66.76999999999997	39.03999999999985	1	6	-0.11	3 TO 6 TIMES A WEEK	PUBLIC	MEDIUM	MEDIUM
16	78	0078005C	FEMALE	68.46999999999997	31.99999999999996	0	6	0.1799999999999997	3 TO 6 TIMES A WEEK	PUBLIC	MEDIUM	HIGH
17	78	0078006C	FEMALE	63.22999999999984	59.15999999999999	1	7	0.1299999999999998	EVERYDAY	PUBLIC	HIGH	MEDIUM
18	78	0078007C	MALE	73.16999999999997	102.19999999999996	1	6	1.68	3 TO 6 TIMES A WEEK	PUBLIC	HIGH	HIGH
19	78	0078009C	MALE	67.16999999999997	55.57999999999984	1	7	1.13	EVERYDAY	PUBLIC	HIGH	HIGH
20	78	0078010C	FEMALE	66.49999999999997	34.38999999999986	1	5	0.6499999999999993	3 TO 6 TIMES A WEEK	PUBLIC	HIGH	MEDIUM
21	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 ακολυθώντας τις εικόνες παρακάτω:



Σχήμα 2: Διαδικασία φόρτωσης δεδομένων στο Jamovi.

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

The screenshot shows the Jamovi software interface. The top menu bar includes 'Variables', 'Data', 'Analyses', 'Plots', and 'Edit'. Below the menu is a toolbar with icons for 'Exploration', 'T-Tests', 'ANOVA', 'Regression', 'Frequencies', 'Factor', and 'Linear Models'. A 'Modules' button is also present. The main area displays a data grid with columns labeled: school\_id, child\_id, child\_gen, child\_age, child\_rea..., child\_kno..., child\_mo..., and socioe. The data consists of 435 rows of child information. At the bottom of the screen, a status bar shows 'Row count 435', 'Filtered 0', 'Deleted 0', 'Added 0', and 'Cells edited 0'. The version 'version 2.7.12' is visible in the bottom right corner.

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

## 1.2 Αρχική επεξεργασία

Όταν εισάγουμε τα δεδομένα μας, το Jamovi κάνει ότι 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 ταμπλάκι.
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 (Σχήμα 4).

	school_id	child_id	child_gen...	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 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 (Σχήμα 5).



DATA VARIABLE

**school\_id**

Description

Measure type: Nominal

Data type: Interval

Missing values

Levels: 97

	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	

Ready Filters 0 Row count 435 Filtered 0 Deleted 0 Added 0 Cells edited 0

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 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.

DATA VARIABLE

**parent\_read\_to\_child**

Description

Measure type: Nominal

Data type: Text

Missing values

Levels: 3 TO 6 TIMES A WEEK, EVERYDAY, NOT AT ALL, ONCE OR TWICE A WEEK

Retain unused levels in analyses

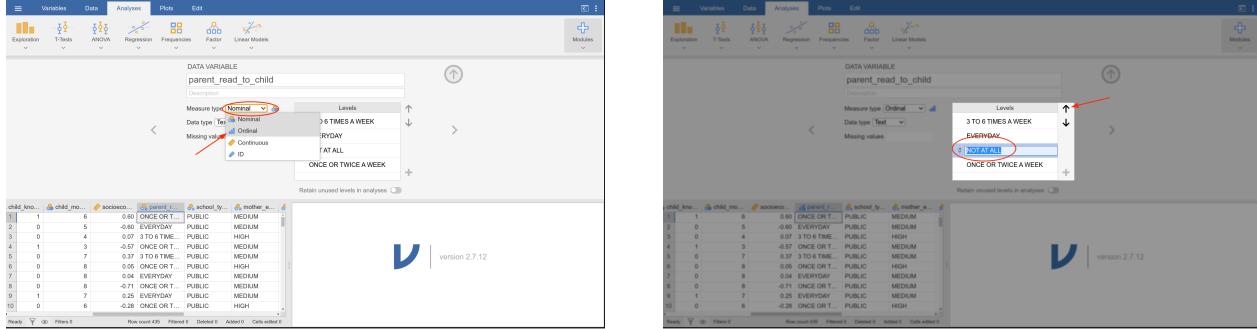
	child_kno...	child_mo...	socioeco...	parent_r...	school_ty...	mother_e...
1	1	6	0.60	ONCE OR T...	PUBLIC	MEDIUM
2	0	5	-0.60	EVERYDAY	PUBLIC	MEDIUM
3	0	4	0.07	3 TO 6 TIME...	PUBLIC	HIGH
4	1	3	-0.57	ONCE OR T...	PUBLIC	MEDIUM
5	0	7	0.37	3 TO 6 TIME...	PUBLIC	MEDIUM
6	0	8	0.05	ONCE OR T...	PUBLIC	HIGH
7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

Ready Filters 0 Row count 435 Filtered 0 Deleted 0 Added 0 Cells edited 0

version 2.7.12

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, βάζομε τις αλλαγές σε αύξουσα σειρά, από τη μικρότερη τιμή στη μεγαλύτερη (Σχήμα 7).

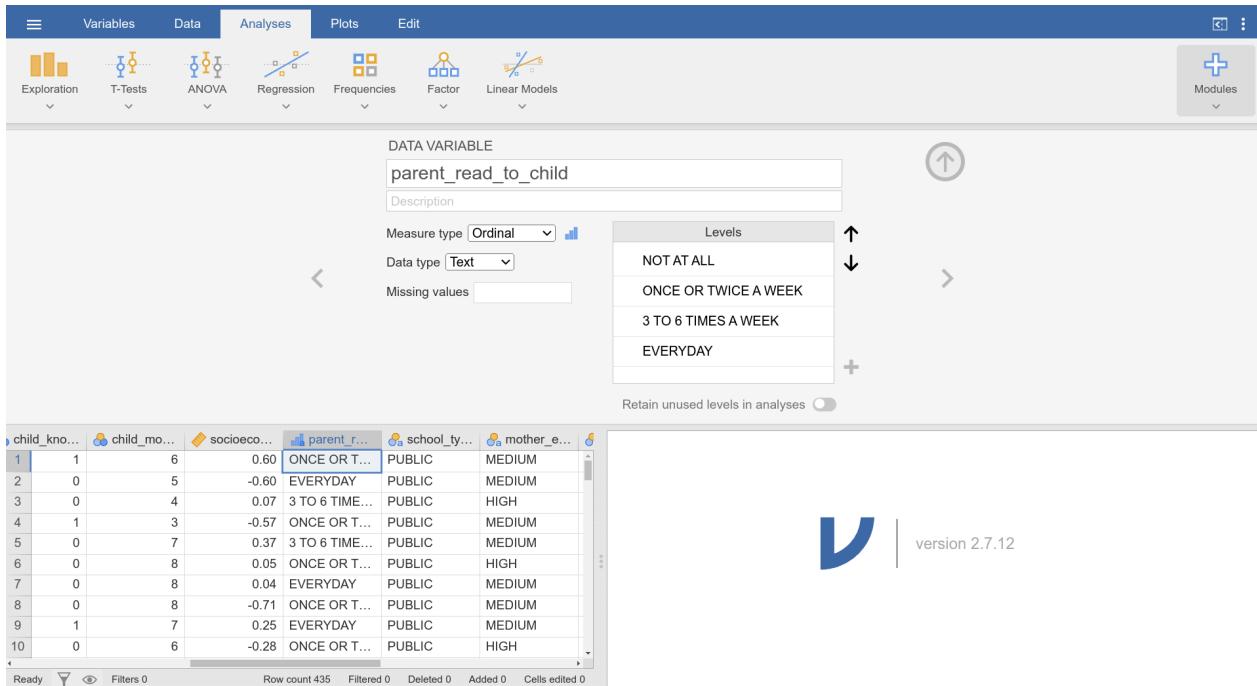


(a) Αλλαγή τύπου της μεταβλητής.

(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.



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

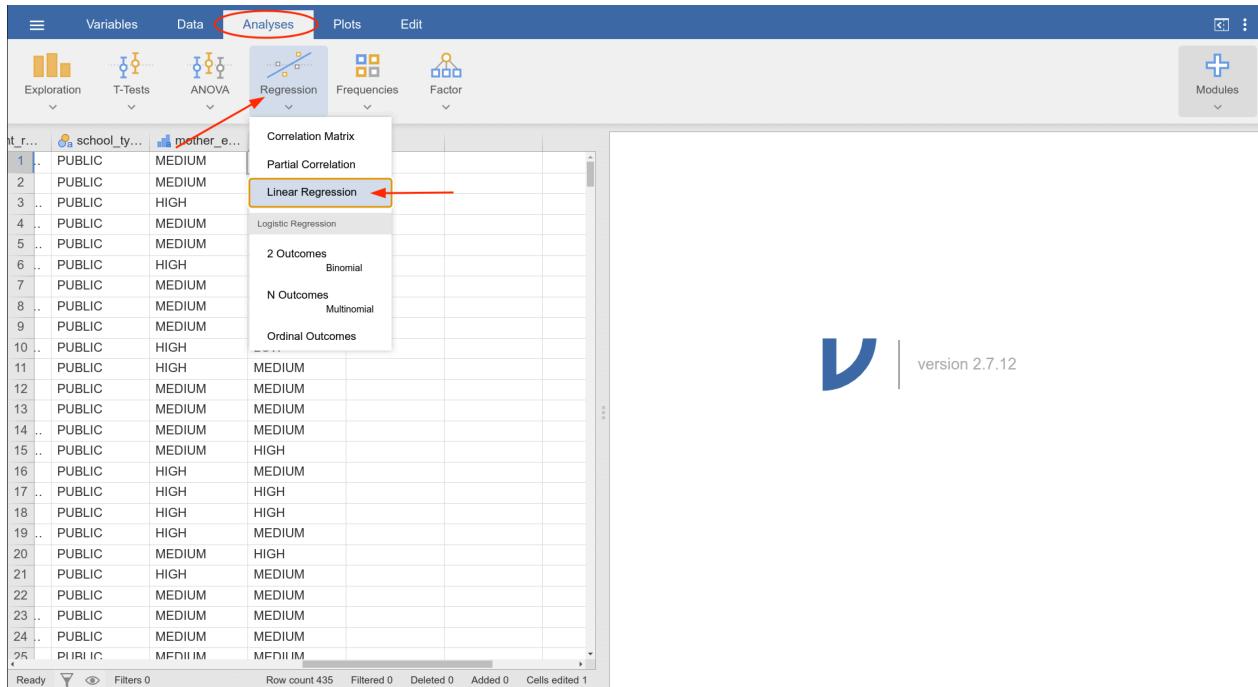
Με τον ίδιο τρόπο αλλάζουμε και τον τύπο των μεταβλητών `mother_education` και `father_education`, οι οποίες πάλι θα έπρεπε να είναι διατάξιμες αλλά έχουν εισαχθεί ως ονομαστικές.

## 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 Jamovi, ακολουθούμε **Analyses > Regression > Linear Regression**.



Σχήμα 9: Η σωστή μορφή που παίρνει η μεταβλητή `parent_read_to_child` μετά τις αλλαγές.

Το Jamovi 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*.

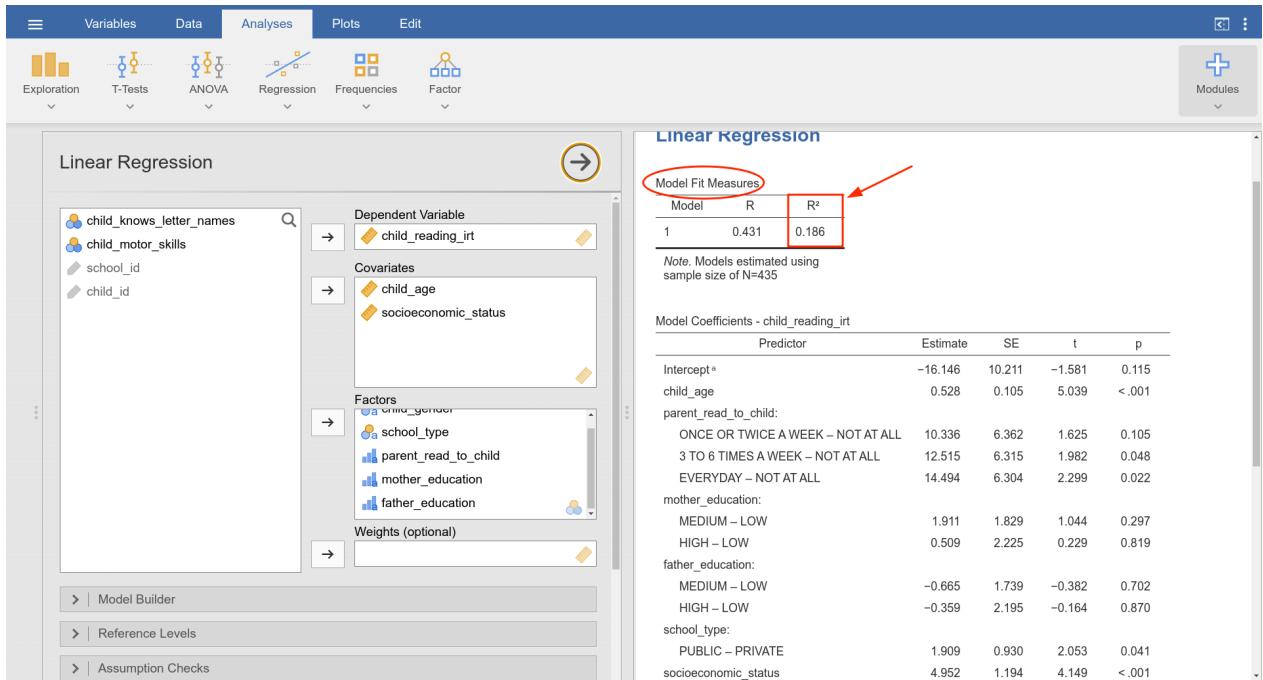
Για να ορίσουμε το μοντέλο της γραμμικής παλινδρόμησης πρέπει σύρουμε και να αφήσουμε τις μεταβλητές που θα χρησιμοποιήσουμε στα σωστά σημεία (Σχήμα 10):

- 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*.

The screenshot shows the Jamovi software interface. At the top, there is a navigation bar with tabs: Variables, Data, Analyses (which is highlighted with a red circle), Plots, and Edit. Below the navigation bar is a toolbar with icons for Exploration, T-Tests, ANOVA, Regression (which is highlighted with a red box), Frequencies, and Factor. The main area contains a data grid with columns labeled 'child\_reading\_irt', 'school\_type', and 'parent\_education'. The data grid has 25 rows of data. To the right of the data grid is a sidebar with various regression models listed: Correlation Matrix, Partial Correlation, Linear Regression (highlighted with a red box and a red arrow pointing to it), Logistic Regression, 2 Outcomes (Binomial), N Outcomes (Multinomial), and Ordinal Outcomes. At the bottom of the interface, there is a status bar with the text 'Ready', 'Row count 435', 'Filtered 0', 'Deleted 0', 'Added 0', and 'Cells edited 1'. On the far right, there is a logo for 'version 2.7.12'.

Σχήμα 10: Ορισμός του μοντέλου γραμμικής παλινδρόμησης.

Στο δεξί παράθυρο του Jamovi 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.

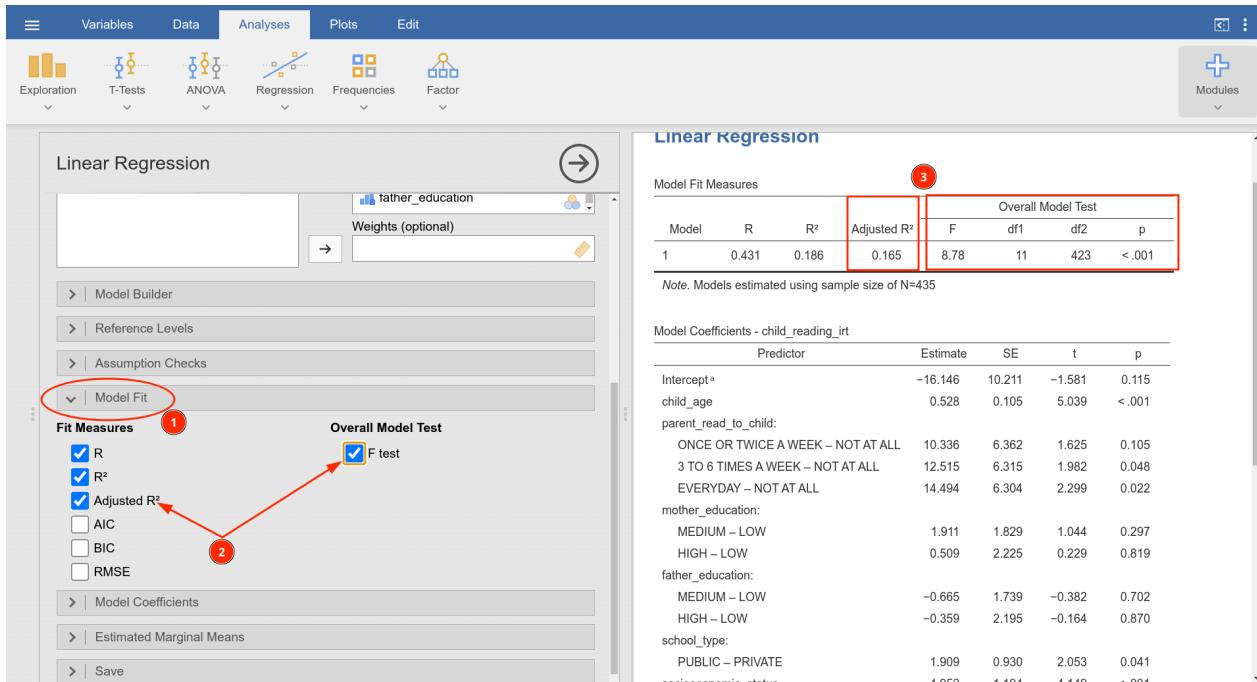


Σχήμα 11: Ορισμός του μοντέλου γραμμικής παλινδρόμησης.

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<sup>2</sup>* (Σχήμα 12). Το προσαρμοσμένο  $R^2$  προσθέτει ένα κόστος στη συμπεριληφθή 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 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  $\beta$  του μοντέλου είναι 0 και να δεχτούμε ότι έστω και ένα από τα  $\beta$  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\_rt*.



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:

$$\begin{aligned} \text{child\_reading\_rt} = & \beta_0 + \beta_1 \times \text{child\_age} + \beta_2 \times \text{socioeconomic\_status} \\ & + \beta_3 \times \text{child\_gender} + \beta_4 \times \text{mother\_education} \\ & + \beta_5 \times \text{father\_education} + \beta_6 \times \text{school\_type} \\ & + \beta_7 \times \text{parent\_read\_to\_child} \end{aligned}$$

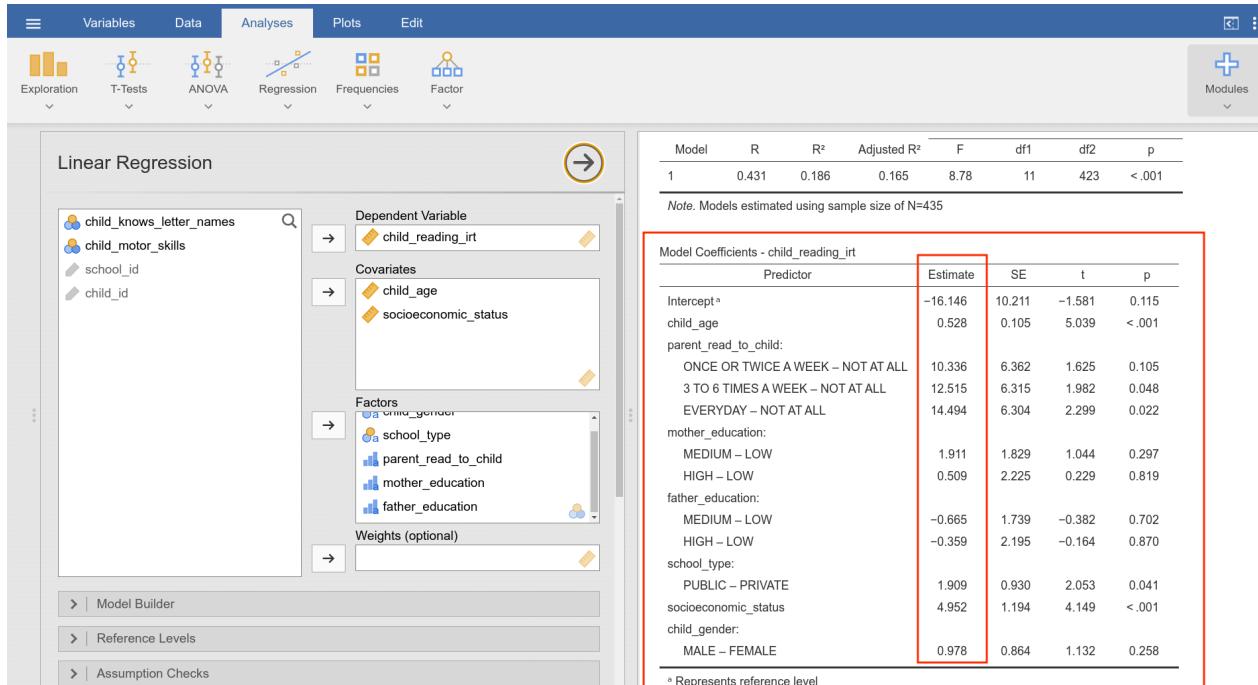
Στο Σχήμα 13, όμως, βλέπουμε πολύ περισσότερα αποτελέσματα από τα 8 που φαίνονται στην παραπάνω εξίσωση. Αυτό συμβαίνει γιατί το Jamovi χωρίζει τις κατηγορικές μεταβλητές σε βοηθητικές μεταβλητές. Για τον χωρισμό σε βοηθητικές μεταβλητές χρησιμοποιείται ένα επίπεδο της κατηγορικής μεταβλητής ως σημείο αναφοράς. Στο Jamovi 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 απορ-

ροφάται από το  $\beta_0$ . Γενικά, όταν έχουμε να κάνουμε με κατηγορικές μεταβλητές, το Jamovi 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.



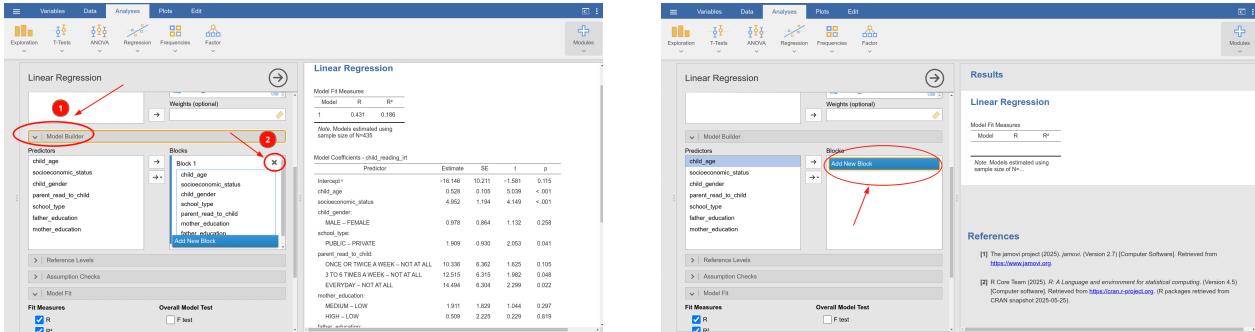
Σχήμα 13: Οι συντελεστές του μοντέλου της γραμμικής παλινδρόμησης που εκτιμήθηκαν χρησιμοποιώντας το δείγμα των 435 μαθητών των δεδομένων μας.

Στο αρχικό μοντέλο συμπεριλάβαμε, μεταξύ άλλων, τις μεταβλητές που περιγράφουν την κοινωνικό/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 κοινωνικό/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  $\beta$ ) είναι 0.

Στο Jamovi, 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 κοινωνικό/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 μορφωτικό επίπεδο των γονέων και το είδος του σχολείου (*mother\_education*, *father\_education*, *school\_type*).

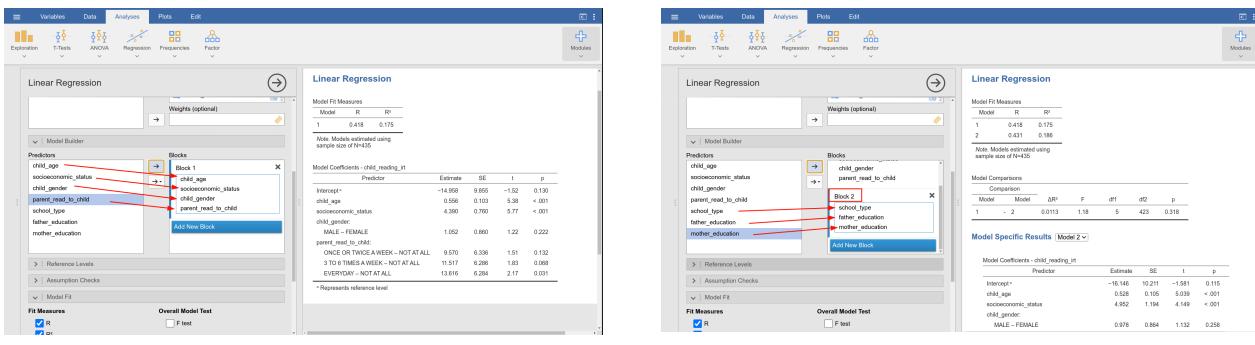
Συνήθως, είναι πιο εύκολο πρώτα να καθαρίσουμε τα Block που υπάρχουν ήδη συμπληρωμένα. Όταν το κάνετε αυτό, τα αποτελέσματα στη δεξιά μεριά της οθόνης θα χαθούν (Σχήμα 14). Στο επόμενο βήμα

Θα συμπληρωθούν ξανά με τη δημιουργία των δύο μοντέλων παλινδρόμησης που έχουμε περιγράψει παραπάνω.



Σχήμα 14: Διαδικασία καθαρισμού υφιστάμενων Block.

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



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.

The screenshot shows the SPSS Statistics interface with the 'Analyses' tab selected. In the center, the 'Linear Regression' dialog box is open, displaying the 'Model Builder' section. On the left, under 'Predictors', variables like 'child\_age', 'socioeconomic\_status', 'child\_gender', 'parent\_read\_to\_child', 'school\_type', 'father\_education', and 'mother\_education' are listed. On the right, under 'Blocks', two blocks are defined: 'Block 1' containing 'child\_gender' and 'parent\_read\_to\_child', and 'Block 2' containing 'school\_type', 'father\_education', and 'mother\_education'. Below the dialog box, the 'Model Fit' section shows 'R' checked and 'F test' unchecked. To the right, the 'Linear Regression' output window displays 'Model Fit Measures' (Model 1: R=0.418, R-squared=0.175; Model 2: R=0.431, R-squared=0.186), a note about sample size (N=435), 'Model Comparisons' (Model 1 vs Model 2: ΔR-squared=0.0113, F=1.18, df1=5, df2=423, p=0.318), and 'Model Specific Results' for Model 2. The 'Model Specific Results' table includes columns for Predictor, Estimate, SE, t, and p.

Model	R	R <sup>2</sup>
1	0.418	0.175
2	0.431	0.186

Comparison	Model	Model	ΔR <sup>2</sup>	F	df1	df2	p
1	-	2	0.0113	1.18	5	423	0.318

Model Coefficients - child_reading_int					
Predictor	Estimate	SE	t	p	
Intercept <sup>a</sup>	-16.146	10.211	-1.581	0.115	
child_age	0.528	0.105	5.039	<.001	
socioeconomic_status	4.952	1.194	4.149	<.001	
child_gender: MALE – FEMALE	0.978	0.864	1.132	0.258	

Σχήμα 16: Τα αποτελέσματα του μερικού *F*-test.

