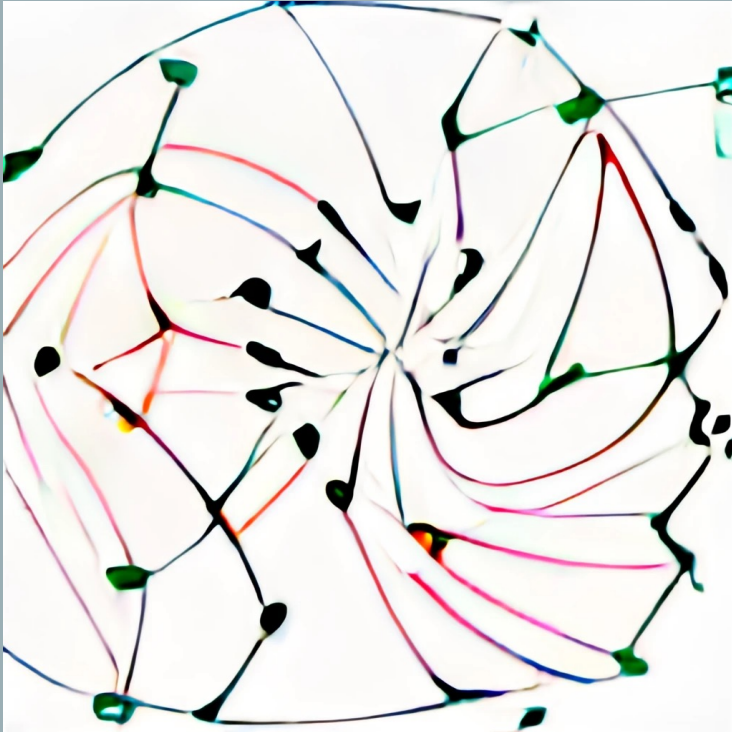


Credit: DALL-E



Άγγελος Μάρκος
amarkos.gr

ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΔΟΜΙΚΩΝ ΕΞΙΣΩΣΕΩΝ

ΜΕ ΤΗ ΧΡΗΣΗ ΤΗΣ R



<https://github.com/amarkos/semworkshop>

Πρόσβαση στο υλικό του σεμιναρίου

ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΔΟΜΙΚΩΝ ΕΞΙΣΩΣΕΩΝ

ΜΕ ΤΗ ΧΡΗΣΗ ΤΗΣ R

ΠΕΡΙΕΧΟΜΕΝΑ

- Η μεγάλη εικόνα
- Τι είναι τα SEM;
- Εφαρμογές των SEM
- Βήματα της μοντελοποίησης
- Λογισμικό

ΜΟΝΤΕΛΟΠΟΙΗΣΗ
ΔΟΜΙΚΩΝ
ΕΞΙΣΩΣΕΩΝ

ΜΕ ΤΗ ΧΡΗΣΗ
ΤΗΣ R

Η ΜΕΓΑΛΗ ΕΙΚΟΝΑ

της στατιστικής μοντελοποίησης

Η ΜΕΓΑΛΗ ΕΙΚΟΝΑ

- Ενδιαφερόμαστε να **κατανοήσουμε** μια πτυχή της **πραγματικότητας**.



ΠΡΑΓΜΑΤΙΚΟΤΗΤΑ

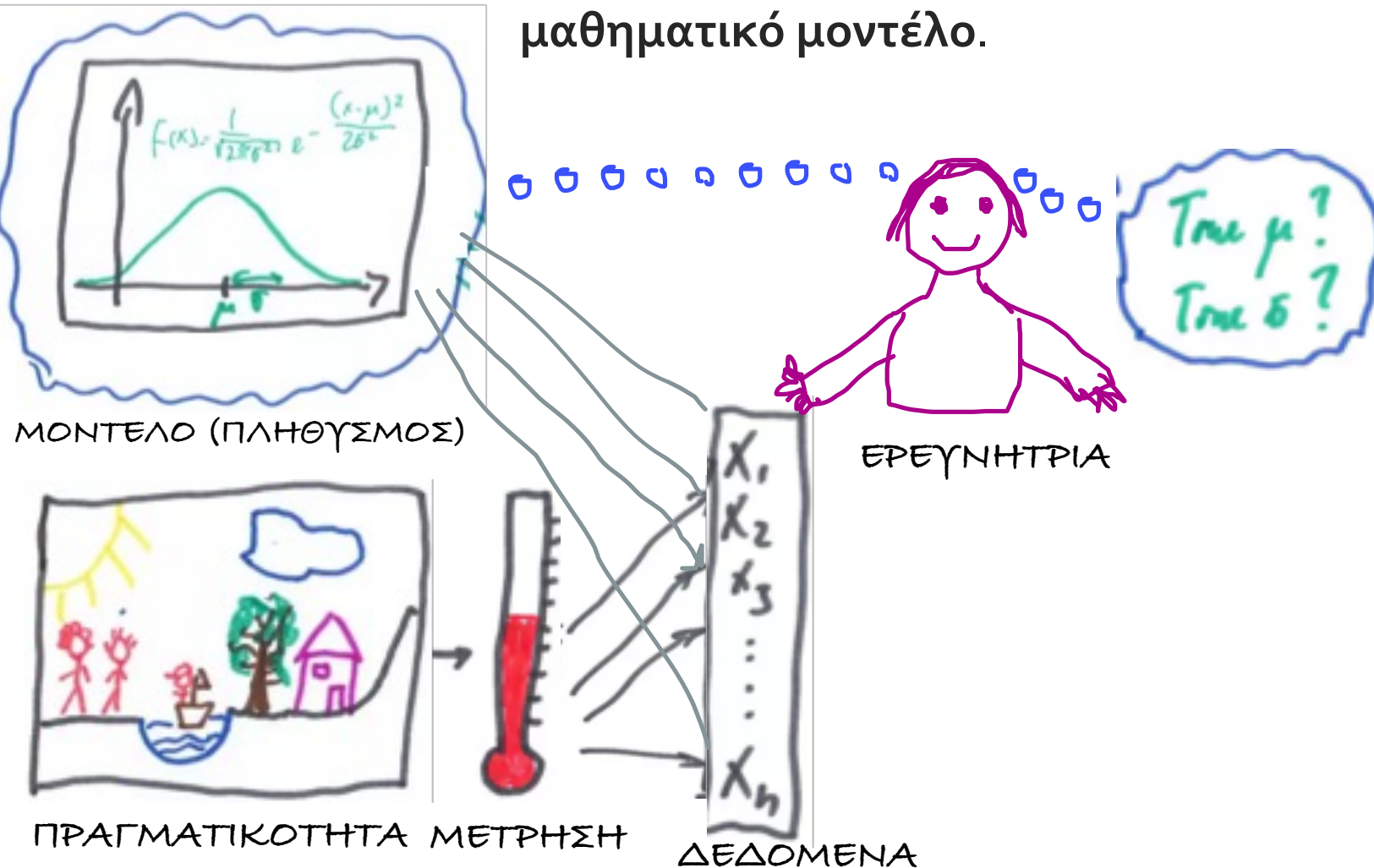
Η ΜΕΓΑΛΗ ΕΙΚΟΝΑ

- Ενδιαφερόμαστε να **κατανοήσουμε** μια πτυχή της **πραγματικότητας** και **συλλέγουμε δεδομένα** (δείγμα) από τον πραγματικό κόσμο.



Η ΜΕΓΑΛΗ ΕΙΚΟΝΑ

- Υποθέτουμε ότι τα δεδομένα μας έχουν γεννηθεί από ένα μαθηματικό μοντέλο.



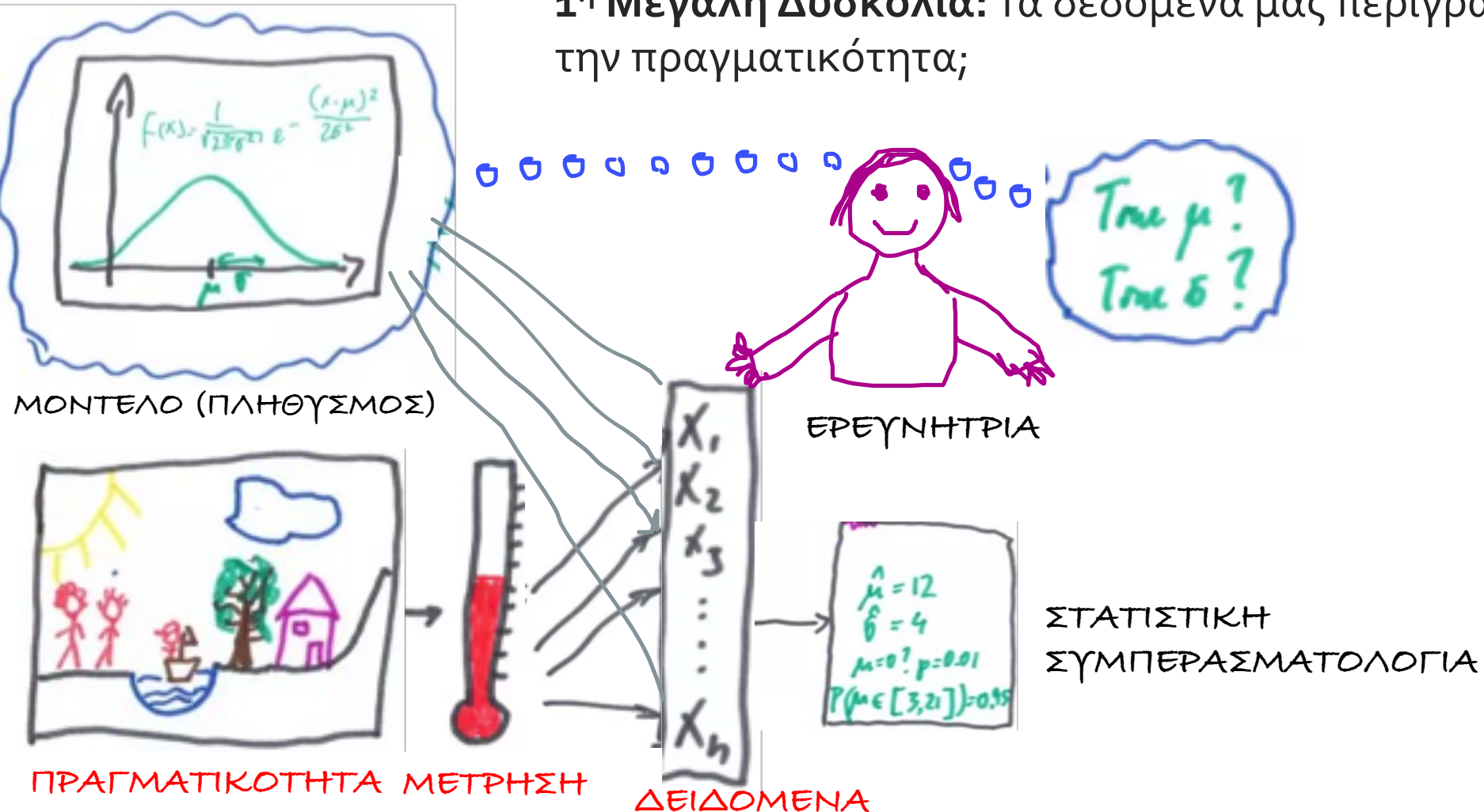
Η ΜΕΓΑΛΗ ΕΙΚΟΝΑ

- Με βάση τις πληροφορίες που παίρνουμε από τα δεδομένα μας, εκτιμούμε τις **παραμέτρους του μοντέλου** και ελέγχουμε πόσο καλά προσαρμόζεται στα δεδομένα.



Η ΜΕΓΑΛΗ ΕΙΚΟΝΑ

1^η Μεγάλη Δυσκολία: Τα δεδομένα μας περιγράφουν καλά την πραγματικότητα;



Η ΜΕΓΑΛΗ ΕΙΚΟΝΑ

2^η Μεγάλη Δυσκολία: Το μαθηματικό μας μοντέλο περιγράφει καλά την πραγματικότητα;



ALL DATASETS ARE BIASED,
BUT SOME ARE USEFUL

Isabelle Guyon



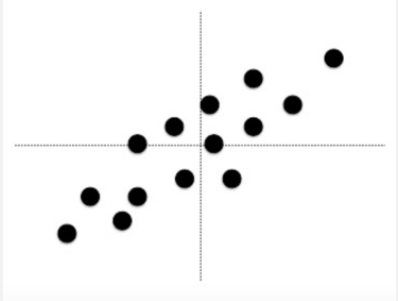
ALL **MODELS ARE WRONG**,
BUT SOME ARE USEFUL

George Box



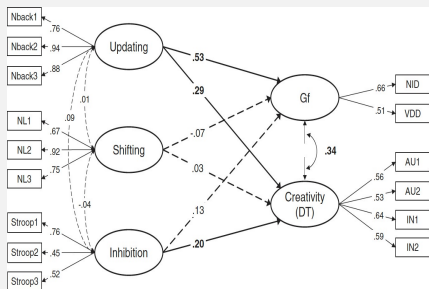
Τι είναι
τα ΜΟΝΤΕΛΑ ΔΟΜΙΚΩΝ ΕΞΙΣΩΣΕΩΝ (SEM)

ΤΙ ΕΙΝΑΙ ΤΑ SEM



συνδιακύμανση
 $\text{cov}(X,Y) = r_{XY}S_XS_Y$

- Τα SEM είναι μια **οικογένεια** στατιστικών μοντέλων τα οποία αναφέρονται σε **σχέσεις** μεταξύ πολλών μεταβλητών. Θεμελιώδεις έννοιες: **(συν)διακύμανση, συσχέτιση**.
- Στην οικογένεια αυτή εντάσσονται στατιστικά μοντέλα, όπως η **παλινδρόμηση** (regression), η **ανάλυση διαδρομών** (path analysis) και η **επιβεβαιωτική ανάλυση παραγόντων** (CFA).
- Ένα μοντέλο SEM μπορεί να συμπεριλάβει τόσο **μετρήσιμες (παρατηρούμενες)** όσο και **λανθάνουσες μεταβλητές**, δηλαδή υποθετικές εννοιολογικές κατασκευές (π.χ. άγχος, νοημοσύνη).
- Μας επιτρέπουν να **ελέγξουμε ένα θεωρητικό μοντέλο** σχετικά τα δεδομένα μας. Το μοντέλο μπορεί να αναπαρασταθεί με ένα **διάγραμμα διαδρομών** (path diagram).



διάγραμμα διαδρομών

ΠΑΡΑΔΕΙΓΜΑΤΑ ΕΡΕΥΝΗΤΙΚΩΝ ΥΠΟΘΕΣΕΩΝ - ΕΡΩΤΗΜΑΤΩΝ ΠΟΥ ΜΠΟΡΟΥΝ ΝΑ ΑΠΑΝΤΗΘΟΥΝ ΜΕ SEM

Μοντέλο
Μεσολάβησης
(mediation)

- ✓ Η **δημιουργικότητα** (creativity) επηρεάζει **άμεσα** την **ταχύτητα επεξεργασίας** (speed processing) και τη **νοημοσύνη** (intelligence), και **άμεσα και έμμεσα** τη **σχολική επίδοση** (academic achievement) (Rindermann & Neubauer, 2004).

Μοντέλο
Ρύθμισης
(moderation)

- ✓ Η **υποστήριξη** από τη **διοίκηση του πανεπιστημίου** **ρυθμίζει** τη σχέση ανάμεσα στην **αυτοαποτελεσματικότητα** στην **έρευνα** και τα **κίνητρα** για **ανάπτυξη** **ερευνητικής δραστηριότητας** (Zhang et al., 2019).

Μοντέλο
Μέτρησης (CFA)

- ✓ Η **δομή** των **ψυχιατρικών διαταραχών** περιγράφεται από έναν **γενικό παράγοντα ψυχοπαθολογίας**; (Caspi et al., 2013)

Μοντέλο
Εξέλιξης (latent growth curve)

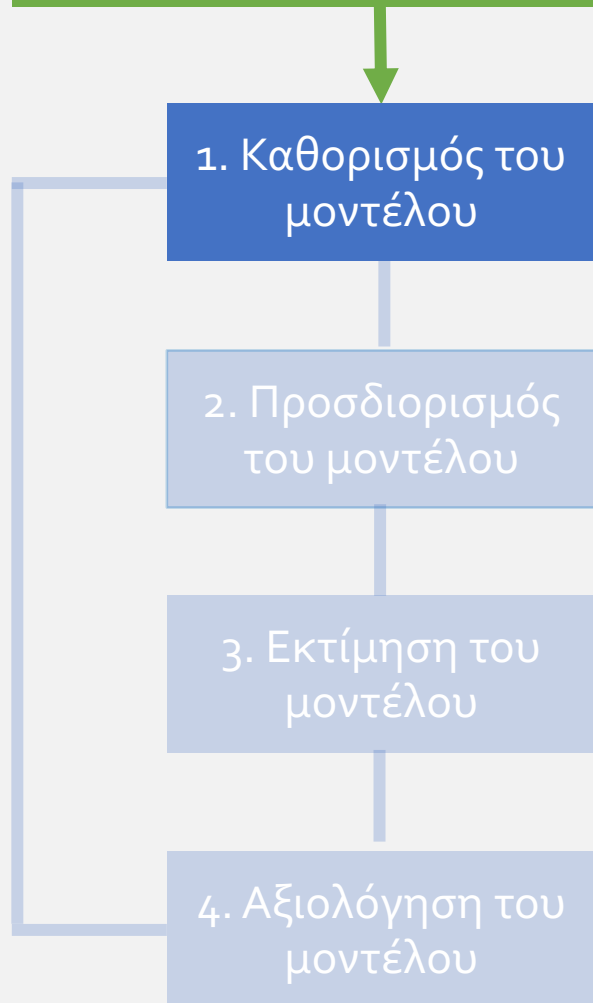
- ✓ Πως **αλλάζουν** οι **απόψεις** για το αν ένα άτομο **συναινεί σε σεξουαλική επαφή** με βάση τα όσα γνωρίζουμε για τη συμπεριφορά του κατά τη διάρκεια μιας **συνεύρεσης**; (Willis & Jozkowski, 2022)

ΤΑ SEM ΣΤΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

- **Επιβεβαιωτικός έλεγχος μιας θεωρίας (theory driven)**
Ωστόσο, **η αιτιότητα δεν προκύπτει από τη στατιστική**, αλλά αντίθετα είναι το μοντέλο SEM που βασίζεται στις θεωρητικές υποθέσεις των ερευνητών περί αιτιότητας!
- **Έλεγχος εναλλακτικών μοντέλων**
Εναλλακτικά μοντέλα SEM μπορεί να προσαρμόζονται **το ίδιο καλά** στα ίδια εμπειρικά δεδομένα...
- **Εντοπισμός του «καλύτερου» μοντέλου που προσαρμόζεται στα δεδομένα (data driven)**
Δεν αποτελεί καλή πρακτική, αλλά δυστυχώς χρησιμοποιείται συχνά στην πράξη χωρίς να αναφέρεται από τους ερευνητές.

Βήματα της ΜΟΝΤΕΛΟΠΟΙΗΣΗΣ ΔΟΜΙΚΩΝ ΕΞΙΣΩΣΕΩΝ

Θεωρία και Ερευνητικές Υποθέσεις



Μεταγραφή των ερευνητικών υποθέσεων στο μοντέλο που επιθυμούμε να ελέγξουμε.

- Ορισμός συντελεστών παλινδρόμησης, διακυμάνσεων, συνδιακυμάνσεων, ...
- Ορισμός παρατηρούμενων και λανθανουσών μεταβλητών.

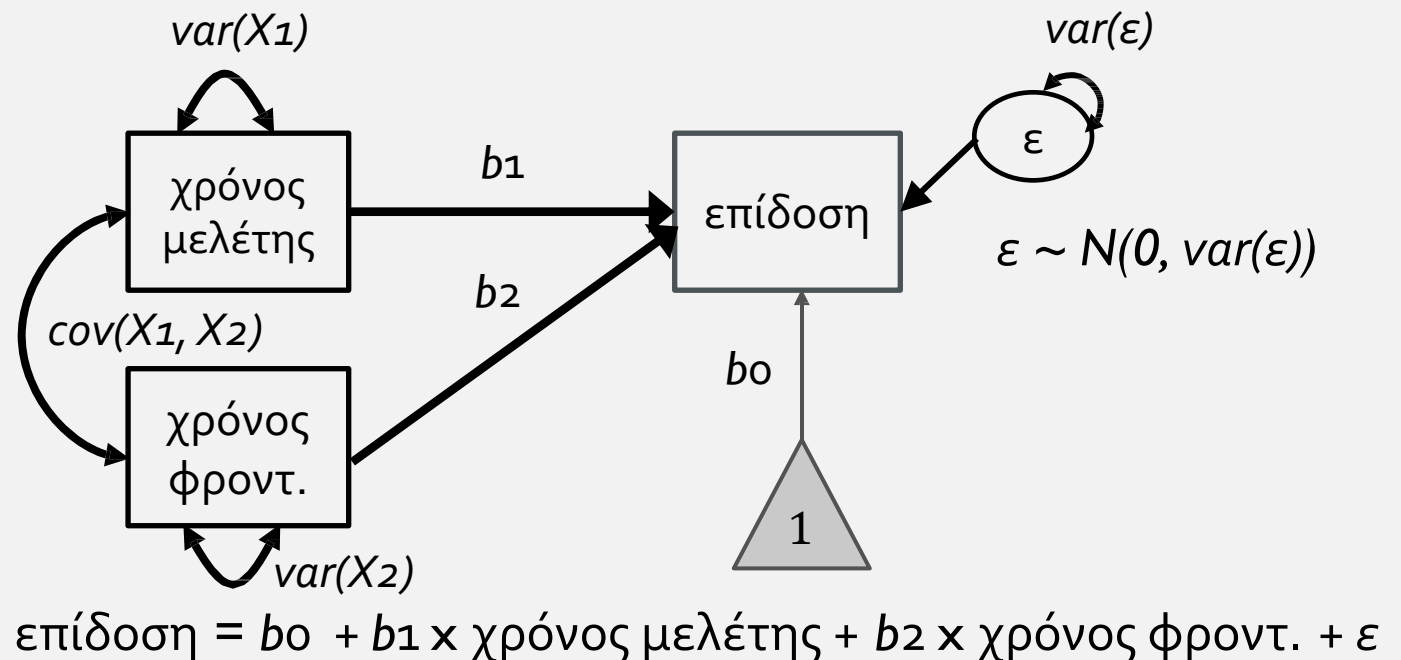
lavaan:

- Παλινδρόμηση: $Y \sim X$
- Συνδιακύμανση: $Y \sim\sim X$
- Λανθάνουσα μεταβλητή: $\text{eta} \sim x1 + x2 + x3$

ΑΠΟ ΤΗΝ ΕΡΕΥΝΗΤΙΚΗ ΥΠΟΘΕΣΗ ΣΤΟ ΜΟΝΤΕΛΟ

Ερευνητική υπόθεση

Ο **χρόνος για μελέτη** και ο **χρόνος για φροντιστήριο** επηρεάζουν την **επίδοση** στις Πανελλαδικές εξετάσεις.



International Studies in Sociology of Education
Vol. 22, No. 2, June 2012, 125–145

 **Routledge**
Taylor & Francis Group

School achievement and family background in Greece: a new exploration of an omnipresent relationship

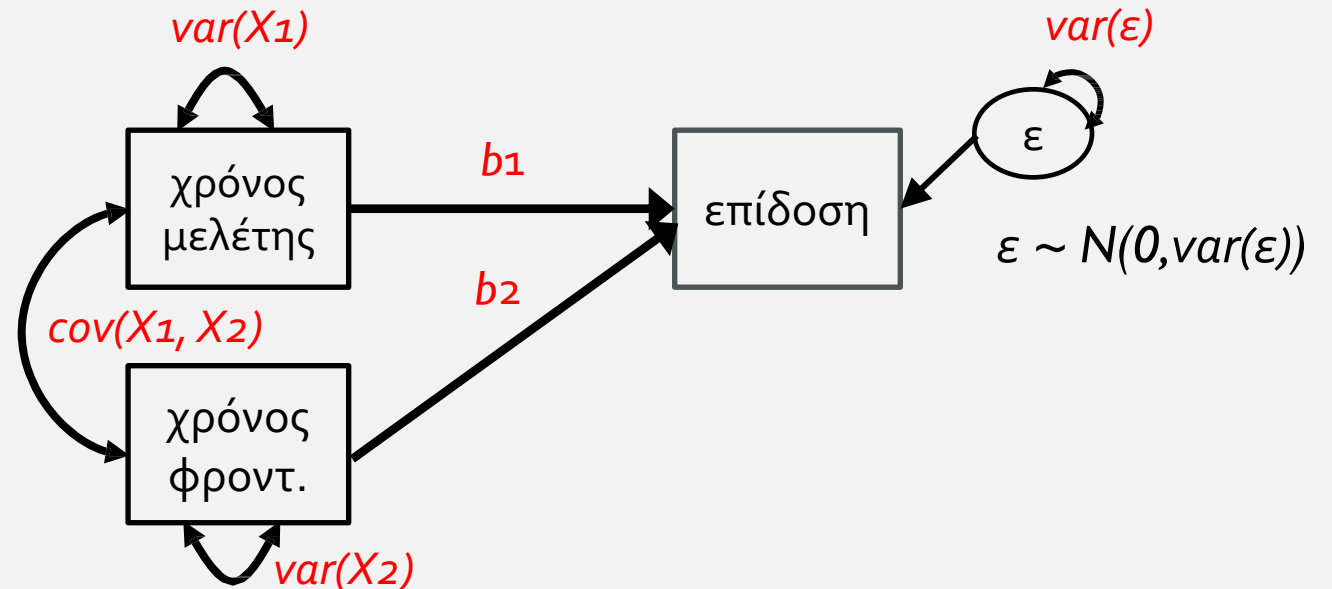
Dionysios Gouvias^{a*}, Athanassios Katsis^b and Aristeia Limakopoulou^c

^aDepartment of the Sciences of Preschool Education and Educational Design, University of the Aegean, Rhodes, Greece; ^bDepartment of Social and Educational Policy, University of Peloponnese, Korinthos, Greece; ^cEducational Research Center, Athens, Greece

(Received 8 September 2011; final version received 16 February 2012)

ΤΟ ΜΟΝΤΕΛΟ

Μοντέλο γραμμικής παλινδρόμησης με δύο ανεξάρτητες μεταβλητές και μία εξαρτημένη.

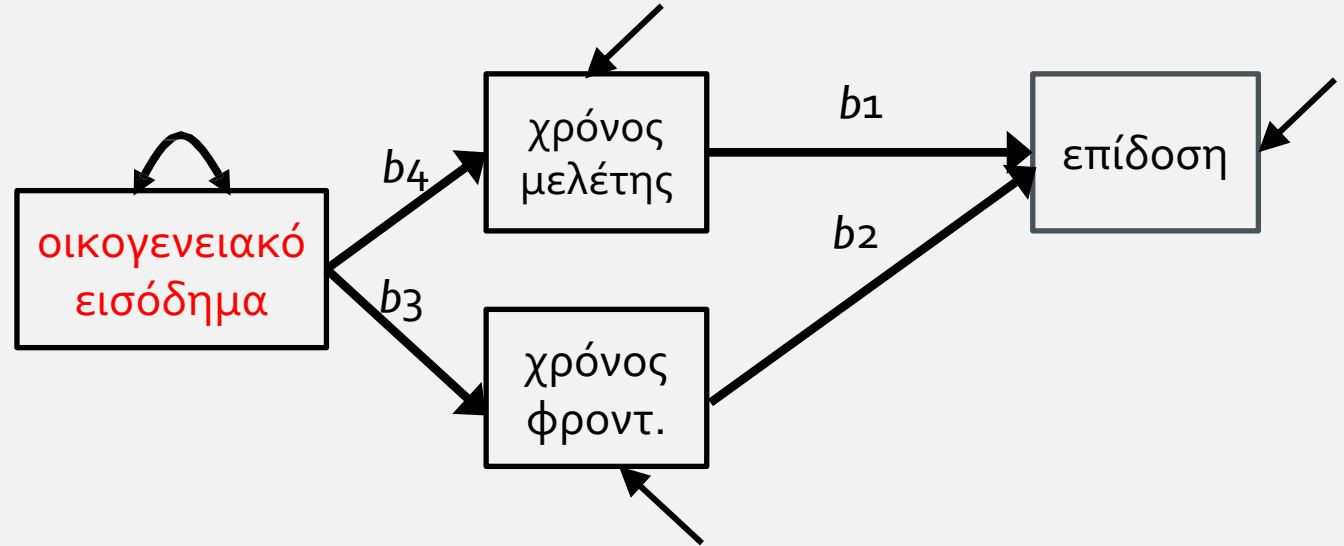


$$\text{επίδοση} = b_1 \times \text{χρόνος μελέτης} + b_2 \times \text{χρόνος φροντ.} + \varepsilon$$

ΑΣ ΠΡΟΣΘΕΣΟΥΜΕ ΜΙΑ ΜΕΤΑΒΛΗΤΗ ΣΤΟ ΜΟΝΤΕΛΟ...

Ερευνητική υπόθεση

Το **οικογενειακό εισόδημα** επηρεάζει έμμεσα την επίδοση στις Πανελλήνιες εξετάσεις μέσω του **χρόνου για μελέτη** και του **χρόνου για φροντιστήριο**.



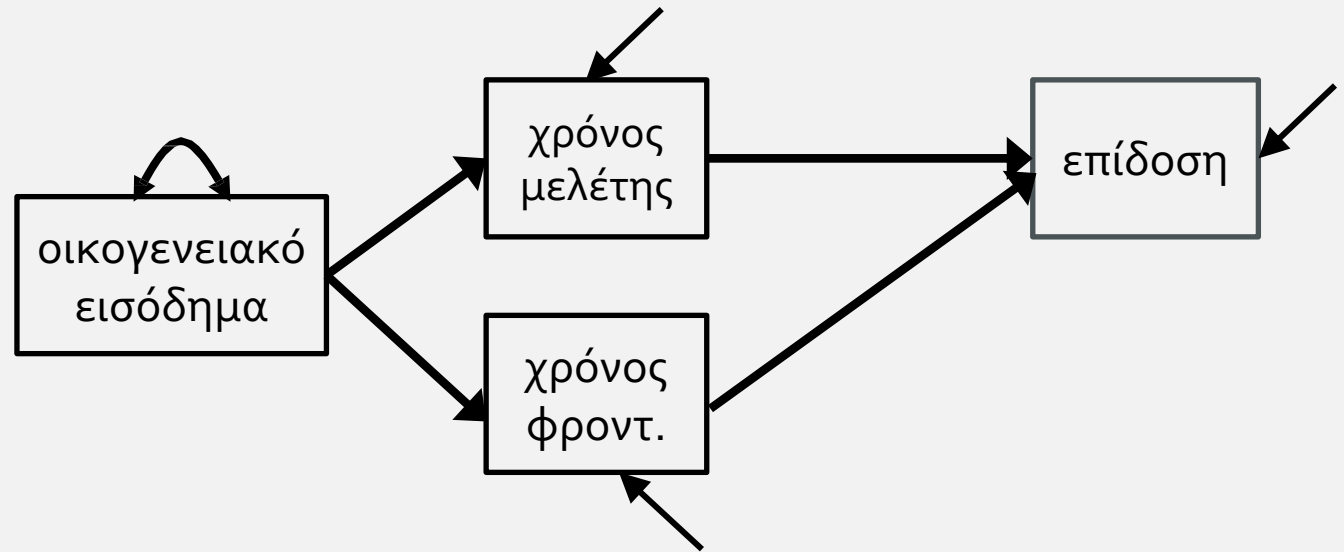
$$\text{επίδοση} = b_1 \times \text{χρόνος μελέτης} + b_2 \times \text{χρόνος φροντ.} + \varepsilon_{\varepsilon}$$

$$\text{χρόνος φροντ.} = b_3 \times \text{οικογενειακό εισόδημα} + \varepsilon_{\chi\varphi}$$

$$\text{χρόνος μελέτης} = b_4 \times \text{οικογενειακό εισόδημα} + \varepsilon_{\chi\mu}$$

ΑΣ ΠΡΟΣΘΕΣΟΥΜΕ ΜΙΑ ΜΕΤΑΒΛΗΤΗ ΣΤΟ ΜΟΝΤΕΛΟ...

Μοντέλο μεσολάβησης με μία ανεξάρτητη μεταβλητή, δύο μεταβλητές μεσολάβησης και μία εξαρτημένη.



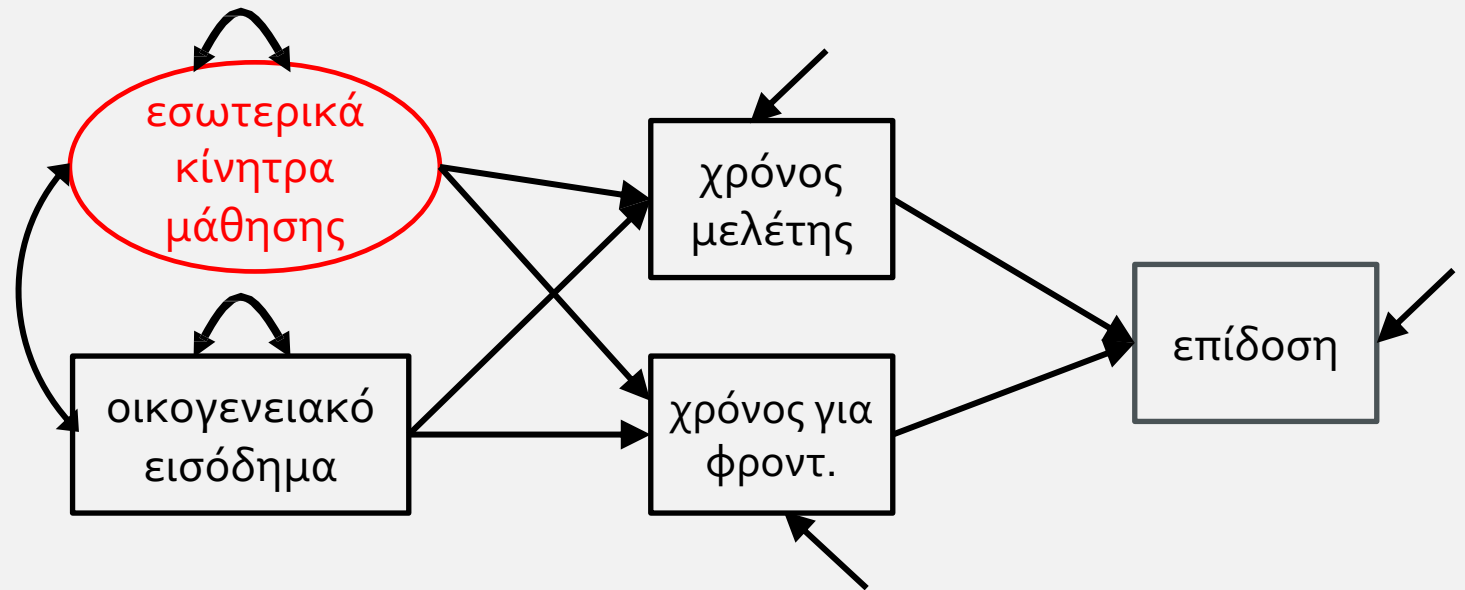
$$\text{επίδοση} = b_1 \times \text{χρόνος μελέτης} + b_2 \times \text{χρόνος φροντ.} + \varepsilon_{\varepsilon}$$

$$\text{χρόνος φροντ.} = b_3 \times \text{οικογενειακό εισόδημα} + \varepsilon_{\chi\varphi}$$

$$\text{χρόνος μελέτης} = b_4 \times \text{οικογενειακό εισόδημα} + \varepsilon_{\chi\mu}$$

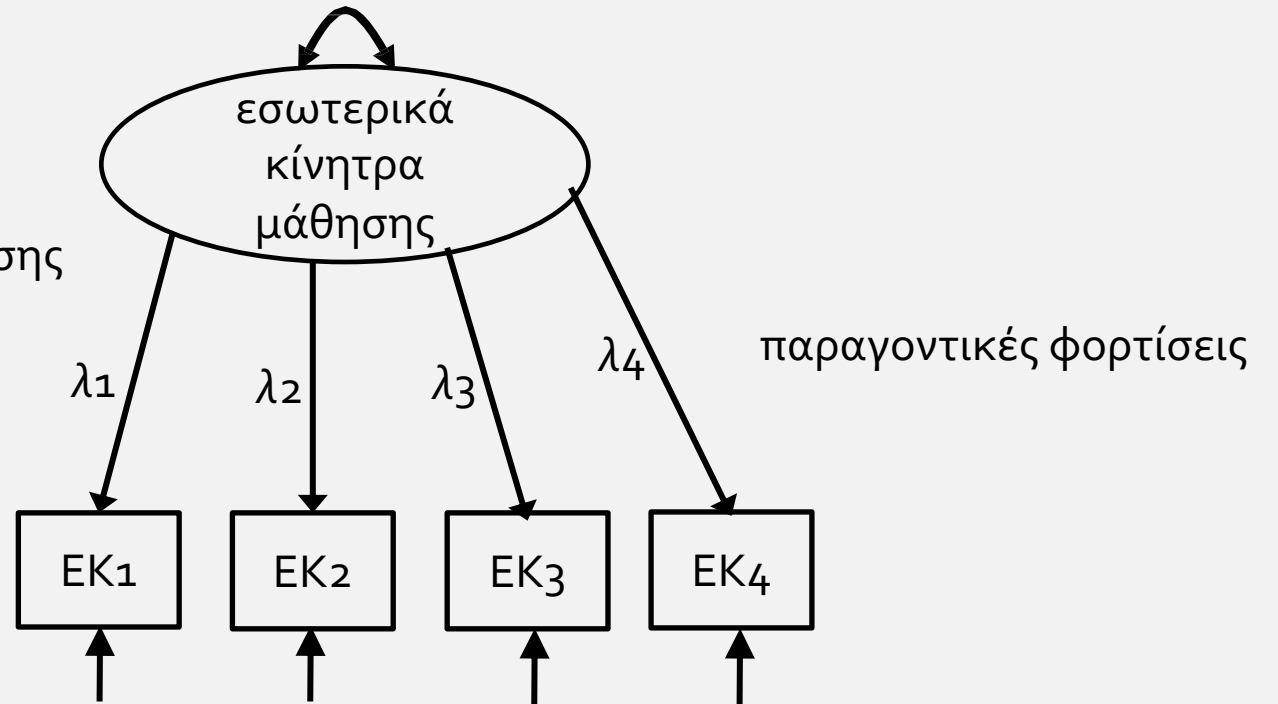
ΑΣ ΠΡΟΣΘΕΣΟΥΜΕ ΑΛΛΗ ΜΙΑ ΜΕΤΑΒΛΗΤΗ ΣΤΟ ΜΟΝΤΕΛΟ...

Μοντέλο με μία εξωγενή λανθάνουσα μεταβλητή, μία εξωγενή παρατηρούμενη μεταβλητή, μία μεταβλητή μεσολάβησης και μία ενδογενή.



Η ΝΕΑ ΜΕΤΑΒΛΗΤΗ ΕΙΝΑΙ ΜΙΑ ΛΑΝΘΑΝΟΥΣΑ ΜΕΤΑΒΛΗΤΗ

Μοντέλο λανθάνουσας μεταβλητής ή μοντέλο μέτρησης
(επιβεβαιωτική ανάλυση παραγόντων – CFA)



EK1: Η φοιτητική ζωή είναι μια από τις καλύτερες περιόδους της ζωής ενός νέου

EK2: Έχω βάλει στόχο να περάσω σε συγκεκριμένο Τμήμα

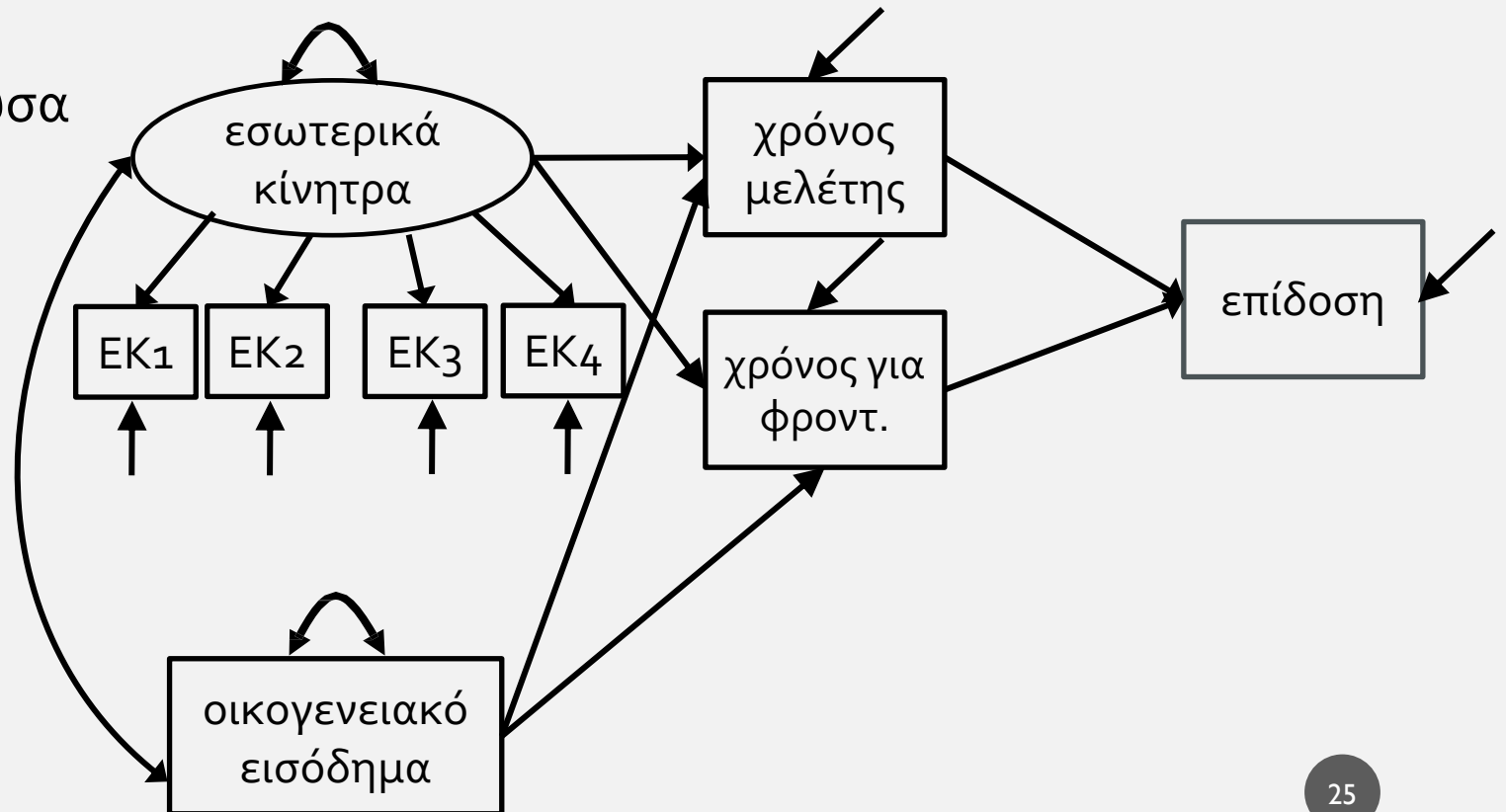
EK3: Πάντα στοχεύω στην υψηλότερη δυνατή ακαδημαϊκή επίδοση

EK4: Το να μαθαίνω κάνει τη ζωή μου πιο ουσιαστική

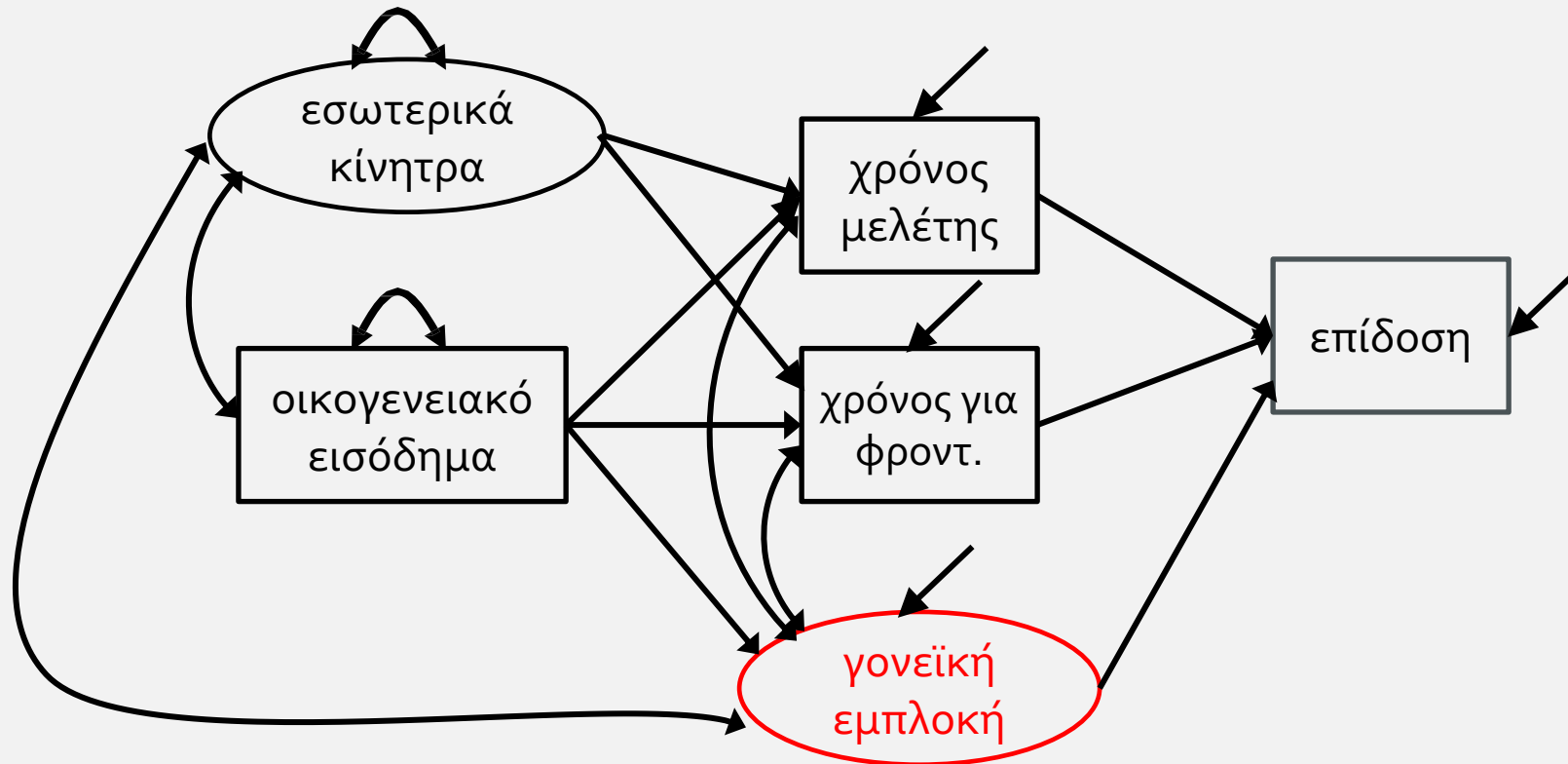
1 = Διαφωνώ απόλυτα
7 = Συμφωνώ απόλυτα

ΤΟ ΜΟΝΤΕΛΟ

Μοντέλο με μία εξωγενή λανθάνουσα μεταβλητή, μία εξωγενή παρατηρούμενη μεταβλητή, μία μεσολαβητική μεταβλητή και μία ενδογενή.

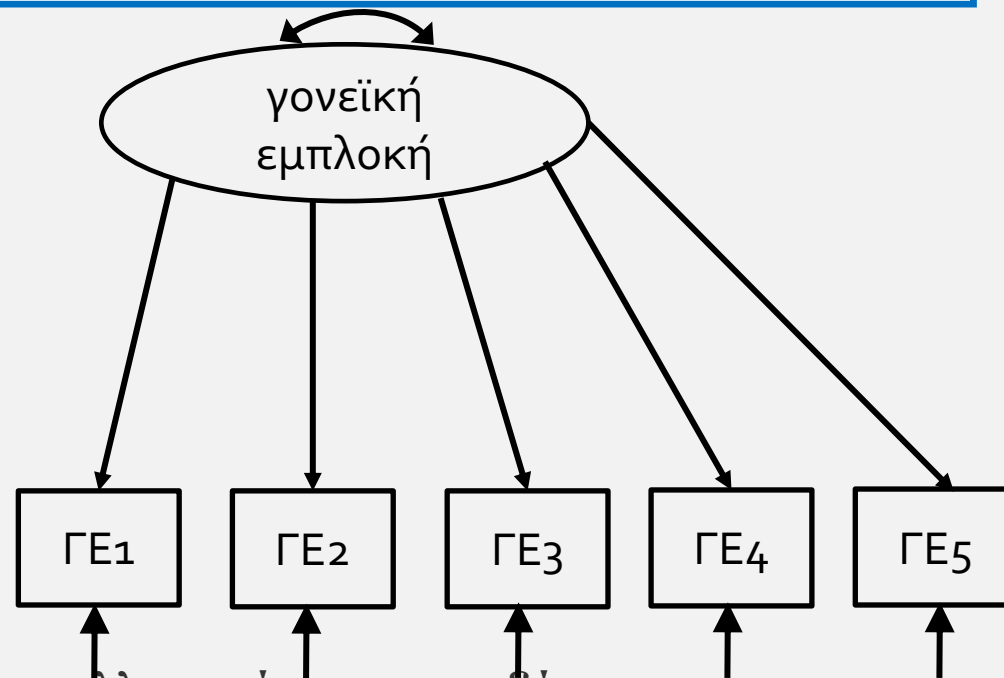


ΑΣ ΠΡΟΣΘΕΣΟΥΜΕ ΑΛΛΗ ΜΙΑ ΜΕΤΑΒΛΗΤΗ ΣΤΟ ΜΟΝΤΕΛΟ...



ΑΛΛΗ ΜΙΑ ΛΑΝΘΑΝΟΥΣΑ ΜΕΤΑΒΛΗΤΗ

Μοντέλο μέτρησης
(επιβεβαιωτική ανάλυση παραγόντων – CFA)



ΓΕ1: Συζητώ συχνά με τους γονείς μου σχετικά με τις μελλοντικές μου σπουδές

ΓΕ2: Οι γονείς μου επικοινωνούν συχνά με τους καθηγητές μου

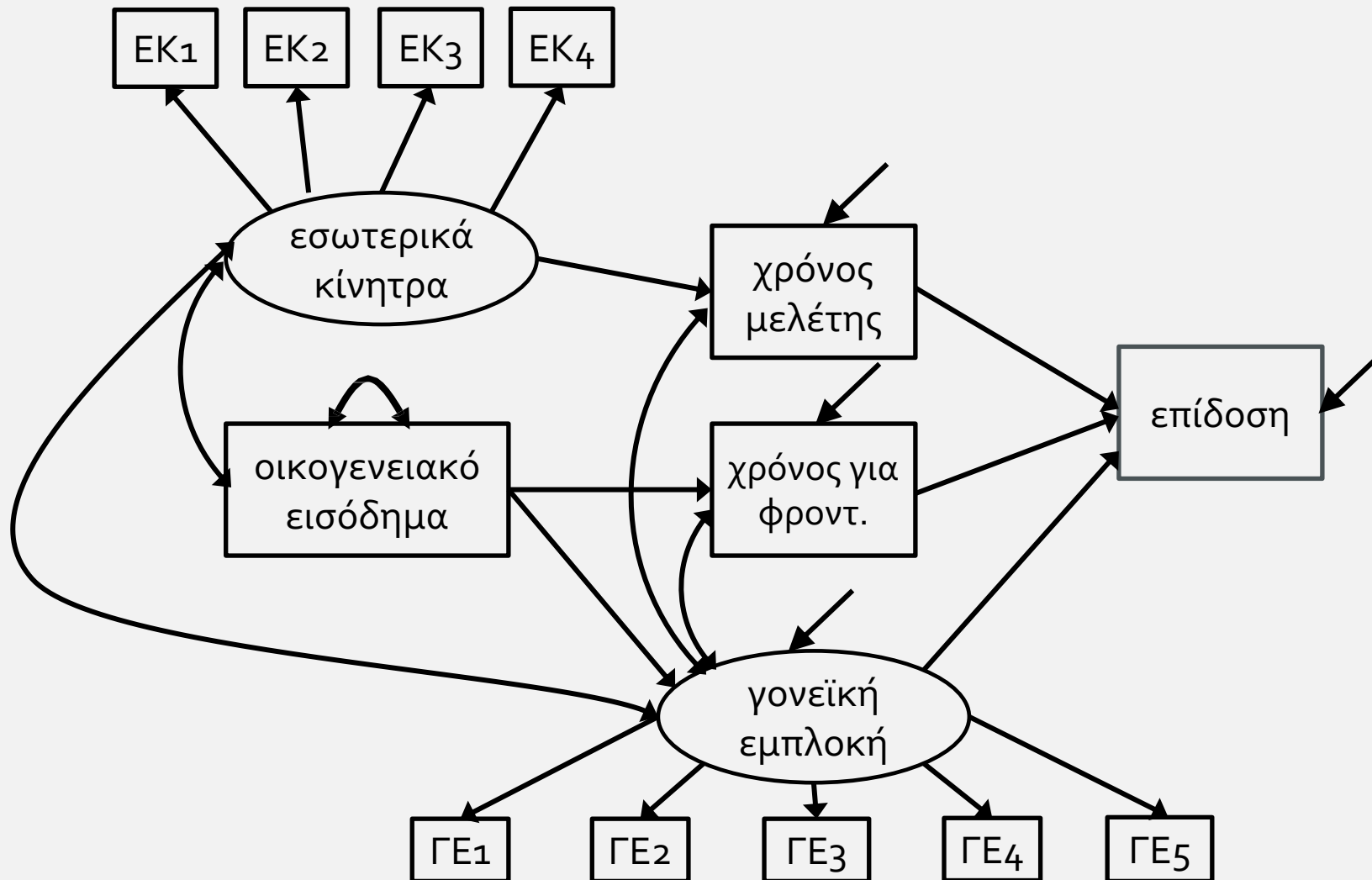
ΓΕ3: Συζητώ συχνά με τους γονείς μου για τα προβλήματα που αντιμετωπίζω με το διάβασμα

ΓΕ4: Οι γονείς μου με βοηθούν όσο μπορούν στις εργασίες που έχω για το σπίτι

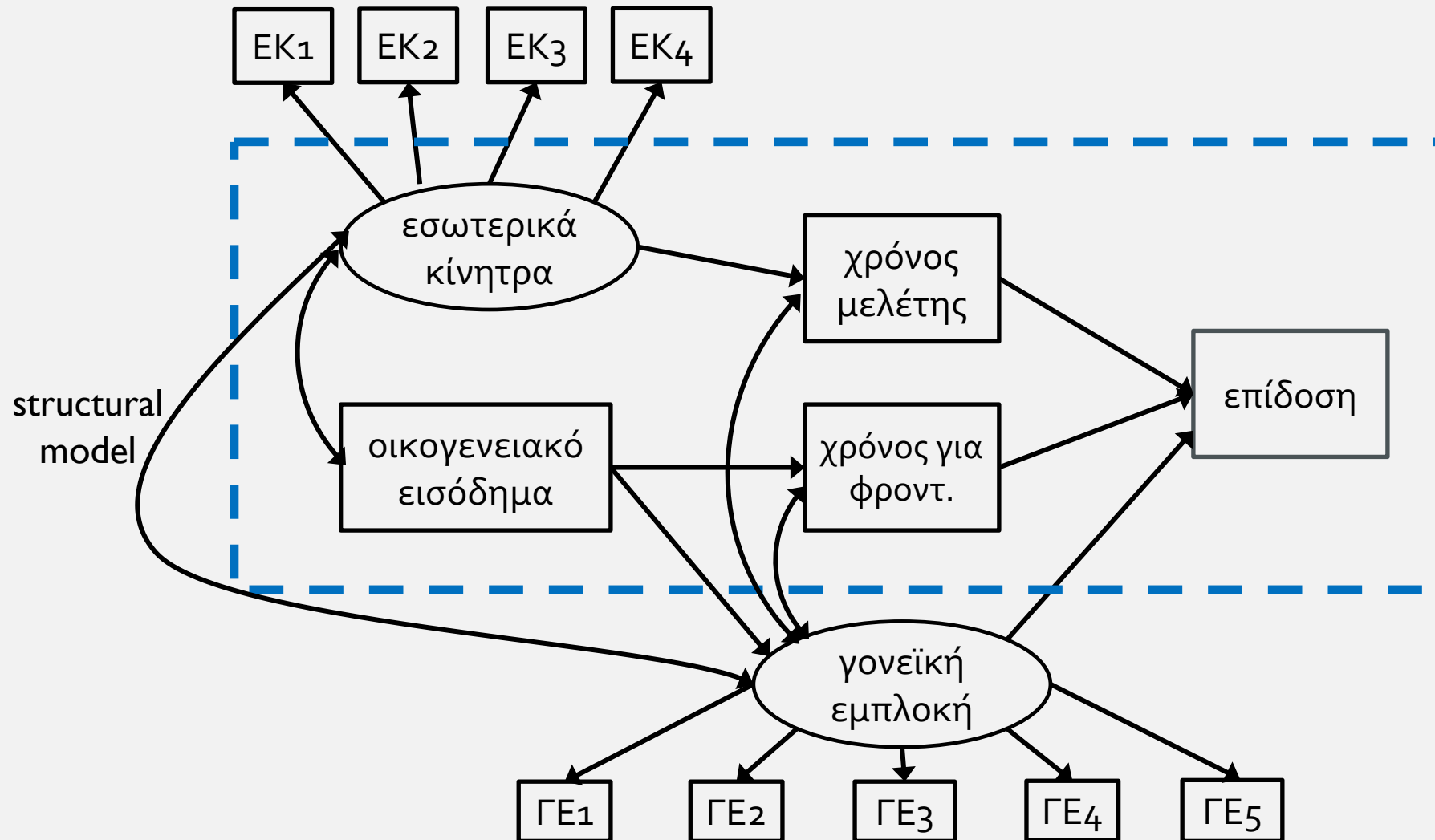
ΓΕ5: Οι γονείς μου φροντίζουν να αποκτήσω τα βοηθητικά βιβλία που θεωρώ απαραίτητα

1 = Διαφωνώ απόλυτα
7 = Συμφωνώ απόλυτα

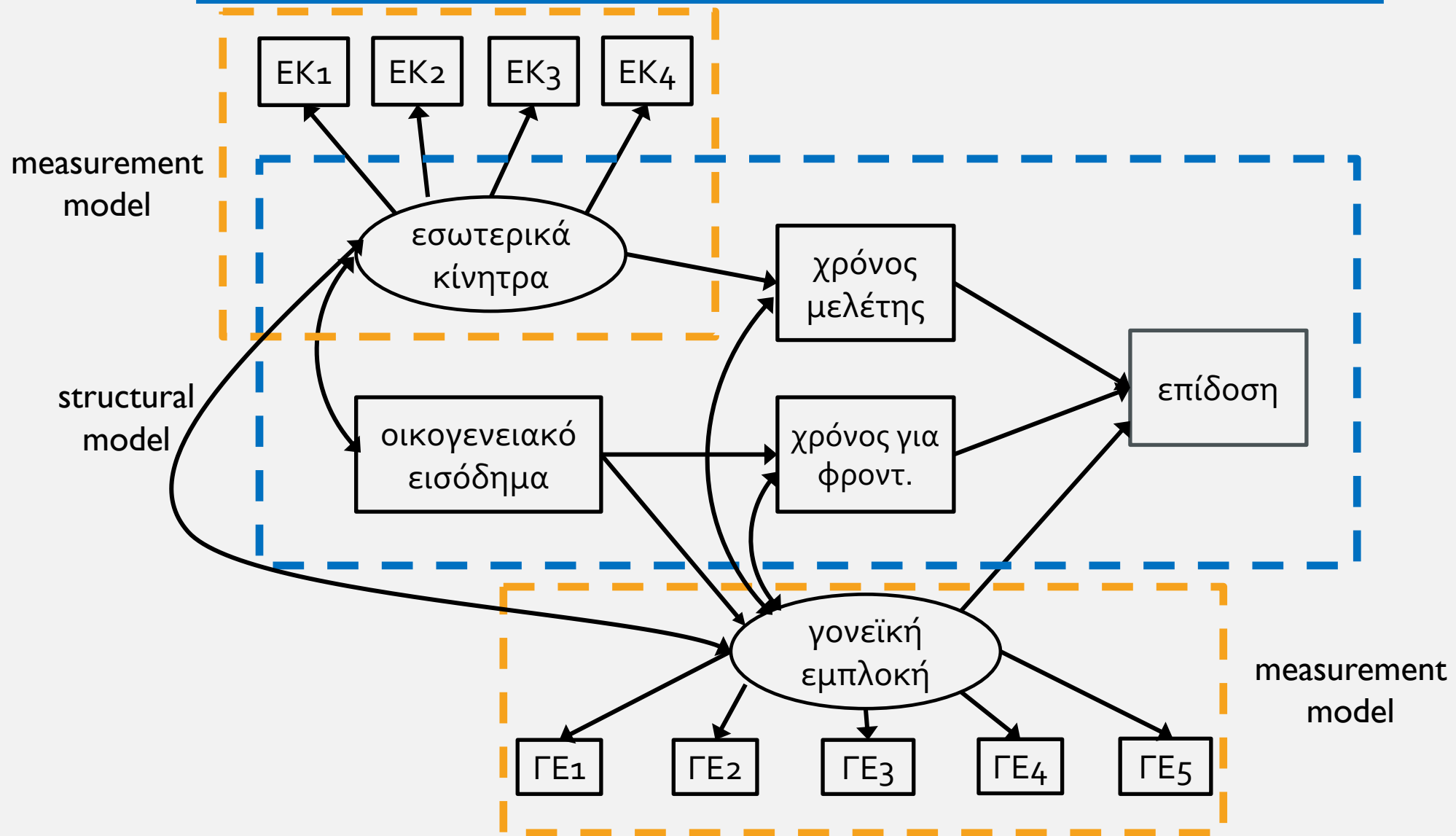
ΤΟ ΤΕΛΙΚΟ ΜΟΝΤΕΛΟ

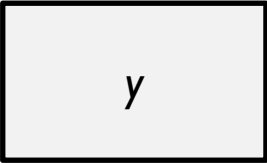
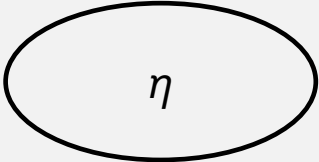


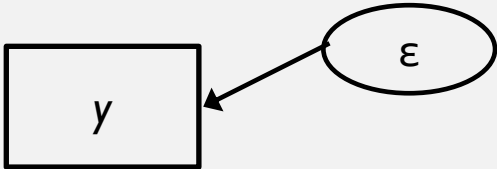
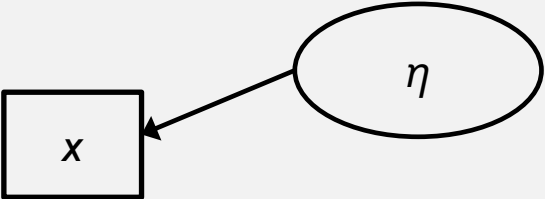
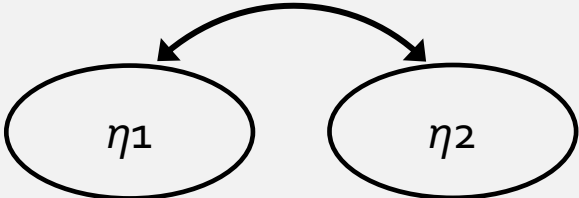


ΔΟΜΙΚΟ ΜΕΡΟΣ ΤΟΥ ΜΟΝΤΕΛΟΥ

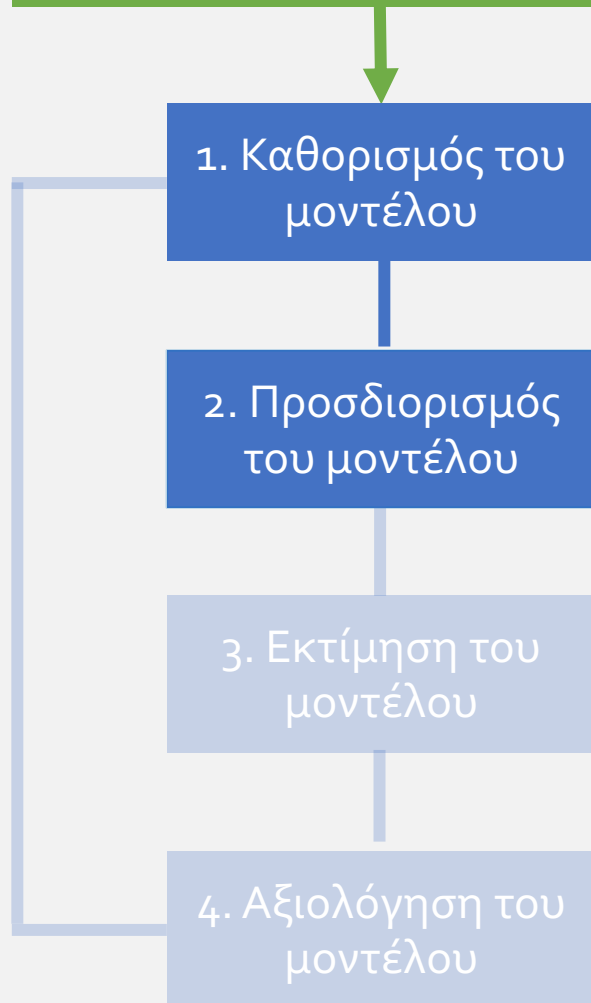


ΜΟΝΤΕΛΑ ΜΕΤΡΗΣΗΣ



Σύμβολο	Περιγραφή
	Παρατηρούμενη ή άμεσα μετρήσιμη μεταβλητή
	Μη Παρατηρούμενη ή λανθάνουσα μεταβλητή
	Άμεση επίδραση/σχέση
	Συσχέτιση ή συνδιακύμανση
	Σφάλμα μέτρησης, ε , της Παρατηρούμενης μεταβλητής y
	Επίδραση της λανθάνουσας μεταβλητής, F στην παρατηρούμενη μεταβλητή, x
	Συσχέτιση μεταξύ λανθανουσών μεταβλητών

Θεωρία και Ερευνητικές Υποθέσεις

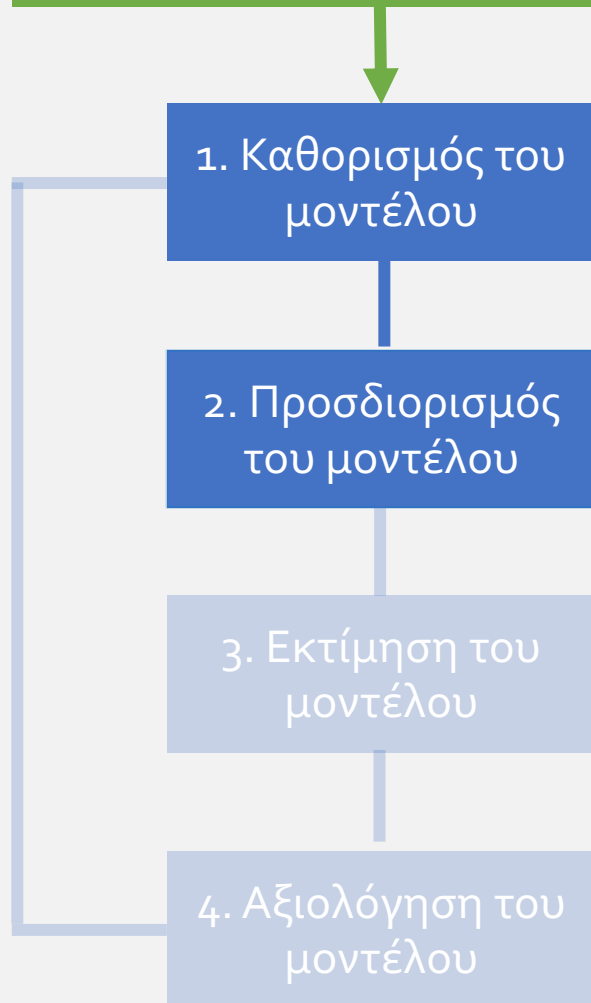


Έχουμε αρκετές πληροφορίες από τα δεδομένα για να εκτιμήσουμε τις παραμέτρους του μοντέλου;

- Πληροφορίες που λαμβάνουμε από τα δεδομένα (διακυμάνσεις, συνδιακυμάνσεις...)
- Ελεύθερα εκτιμώμενες παράμετροι του μοντέλου: διακυμάνσεις, συνδιακυμάνσεις, συντελεστές παλινδρόμησης, σταθεροί όροι, φορτίσεις

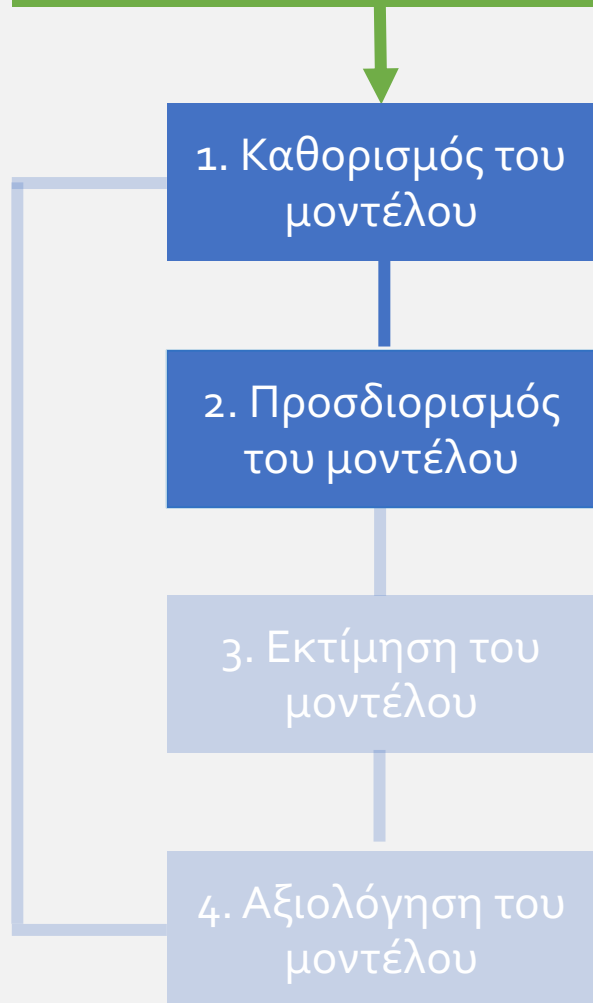
Ο αριθμός των παρατηρούμενων διακυμάνσεων, συνδιακυμάνσεων πρέπει να είναι τουλάχιστον ίσος με τον αριθμό των ελεύθερα εκτιμώμενων παραμέτρων του μοντέλου
→ βαθμοί ελευθερίας του μοντέλου ≥ 0

Θεωρία και Ερευνητικές Υποθέσεις



- ✓ Για να εκτιμηθούν οι παράμετροι ενός μοντέλου είναι απαραίτητο να μπορεί να βρεθεί μια μοναδική λύση για τις τιμές των παραμέτρων.
- ✓ Εάν για όλες τις παραμέτρους του μοντέλου μπορεί να βρεθεί μία και μοναδική λύση τότε το μοντέλο ονομάζεται **υπερ-προσδιορισμένο** (over-identified) ή **προσδιορισμένο** (just-identified).
- ✓ Στην περίπτωση που έστω και μία παράμετρος δεν μπορεί να προσδιοριστεί, τότε το μοντέλο δεν έχει προσδιοριστεί επαρκώς (**υπο-προσδιορισμένο** ή under-identified).

Θεωρία και Ερευνητικές Υποθέσεις

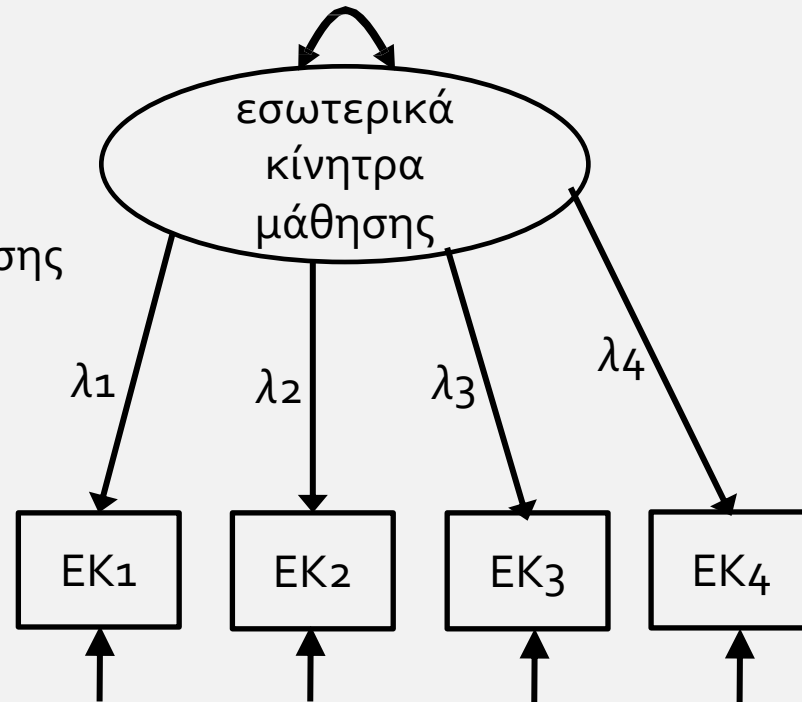


Πως μπορούμε να ελέγξουμε στην πράξη αν το μοντέλο μας έχει προσδιοριστεί (Kline, 2015):

- βαθμοί ελευθερίας ≥ 0
- αναδρομικά (recursive) μοντέλα μπορούν πάντα να προσδιοριστούν (δλδ μοντέλα χωρίς «κύκλους»)
- δίνουμε σημασία στις προειδοποιήσεις (warnings) των λογισμικών
- πολύ μεγάλα τυπικά σφάλματα ή τυπικά σφάλματα που δεν εκτιμήθηκαν → το μοντέλο δεν προσδιορίστηκε
- επανακτίμηση του μοντέλου με διαφορετικές αρχικές τιμές (starting values) πρέπει να οδηγήσει στις ίδιες τιμές εκτιμώμενων παραμέτρων.

ΠΑΡΑΔΕΙΓΜΑ ΠΡΟΣΔΙΟΡΙΣΜΟΥ ΕΝΟΣ ΜΟΝΤΕΛΟΥ ΜΕΤΡΗΣΗΣ

Μοντέλο λανθάνουσας μεταβλητής ή μοντέλο μέτρησης (επιβεβαιωτική ανάλυση παραγόντων – CFA)



EK1: Η φοιτητική ζωή είναι μια από τις καλύτερες περιόδους της ζωής ενός νέου

EK2: Έχω βάλει στόχο να περάσω σε συγκεκριμένο Τμήμα

EK3: Πάντα στοχεύω στην υψηλότερη δυνατή ακαδημαϊκή επίδοση

EK4: Το να μαθαίνω κάνει τη ζωή μου πιο ουσιαστική

1 = Διαφωνώ απόλυτα
7 = Συμφωνώ απόλυτα

ΤΑ ΑΡΧΙΚΑ ΔΕΔΟΜΕΝΑ...

EK1	EK2	EK3	EK4
1	2	2	1
5	4	4	4
4	4	3	3
6	5	6	7
5	5	5	5
5	4	2	4
5	2	4	4
2	2	2	2
3	3	4	3
3	4	4	4
3	2	3	3
4	4	4	3
1	2	1	1
2	2	3	1

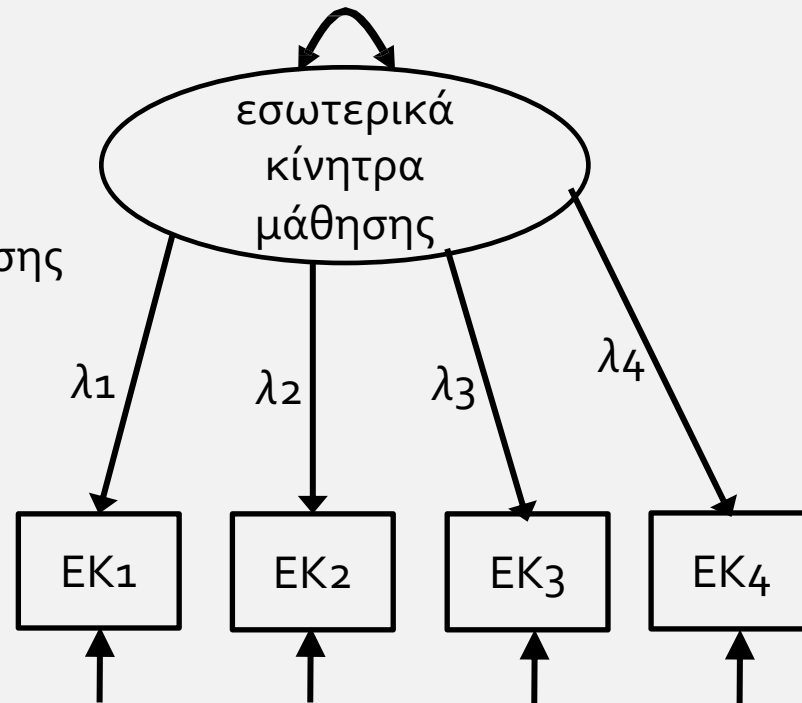
...ΑΝΤΙΣΤΟΙΧΟΥΝ ΣΕ ΕΝΑΝ ΕΜΠΕΙΡΙΚΟ
ΠΙΝΑΚΑ ΔΙΑΚΥΜΑΝΣΗΣ - ΣΥΝΔΙΑΚΥΜΑΝΣΗΣ

	EK1	EK2	EK3	EK4
EK1	4.667	1.542	2.458	1.375
EK2	1.542	3.523	1.965	.583
EK3	2.458	1.965	3.657	1.375
EK4	1.375	.583	1.375	2.583

Από τα δεδομένα μας λαμβάνουμε 10 πληροφορίες,
4 διακυμάνσεις και 6 συνδιακυμάνσεις.

ΤΟ ΜΟΝΤΕΛΟ ΜΑΣ...

Μοντέλο λανθάνουσας μεταβλητής ή μοντέλο μέτρησης
(επιβεβαιωτική ανάλυση παραγόντων – CFA)



ΕΚ1: Η φοιτητική ζωή είναι μια από τις καλύτερες περιόδους της ζωής ενός νέου

ΕΚ2: Έχω βάλει στόχο να περάσω σε συγκεκριμένο Τμήμα

ΕΚ3: Πάντα στοχεύω στην υψηλότερη δυνατή ακαδημαϊκή επίδοση

ΕΚ4: Το να μαθαίνω κάνει τη ζωή μου πιο ουσιαστική

1 = Διαφωνώ απόλυτα
7 = Συμφωνώ απόλυτα

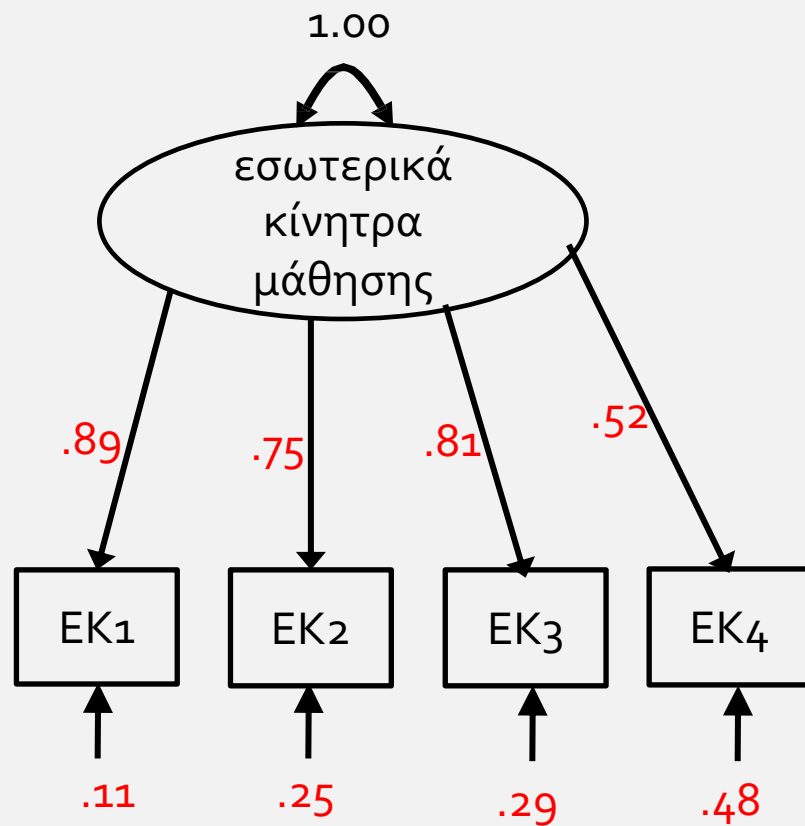
...ΣΥΝΕΠΑΓΕΤΑΙ ΕΝΑΝ ΘΕΩΡΗΤΙΚΟ ΠΙΝΑΚΑ ΔΙΑΚΥΜΑΝΣΗΣ - ΣΥΝΔΙΑΚΥΜΑΝΣΗΣ...

	x1	x2	x3	x4
x1	$\lambda_1^2 + \text{var}(\epsilon_1)$	$\lambda_1 \cdot \lambda_2$	$\lambda_1 \cdot \lambda_3$	$\lambda_1 \cdot \lambda_4$
x2	$\lambda_1 \cdot \lambda_2$	$\lambda_2^2 + \text{var}(\epsilon_2)$	$\lambda_2 \cdot \lambda_3$	$\lambda_2 \cdot \lambda_4$
x3	$\lambda_1 \cdot \lambda_3$	$\lambda_2 \cdot \lambda_3$	$\lambda_3^2 + \text{var}(\epsilon_3)$	$\lambda_3 \cdot \lambda_4$
x4	$\lambda_1 \cdot \lambda_4$	$\lambda_2 \cdot \lambda_4$	$\lambda_3 \cdot \lambda_4$	$\lambda_4^2 + \text{var}(\epsilon_4)$

Οι ελεύθερα εκτιμώμενες παράμετροι του μοντέλου μας είναι 8.

Άρα έχουμε **2 βαθμούς ελευθερίας** (over-identified).

...ΒΑΣΕΙ ΤΟΥ ΟΠΟΙΟΥ ΕΚΤΙΜΩΝΤΑΙ ΟΙ
ΕΛΕΥΘΕΡΕΣ ΠΑΡΑΜΕΤΡΟΙ ΤΟΥ ΜΟΝΤΕΛΟΥ...



...ΕΤΣΙ ΩΣΤΕ ΟΙ ΔΙΑΦΟΡΕΣ ΤΩΝ ΔΥΟ ΠΙΝΑΚΩΝ
ΝΑ ΕΙΝΑΙ ΟΣΟ ΤΟ ΔΥΝΑΤΟ ΜΙΚΡΟΤΕΡΕΣ.

Εμπειρικός πίνακας
διακύμανσης -
συνδιακύμανσης, **S**

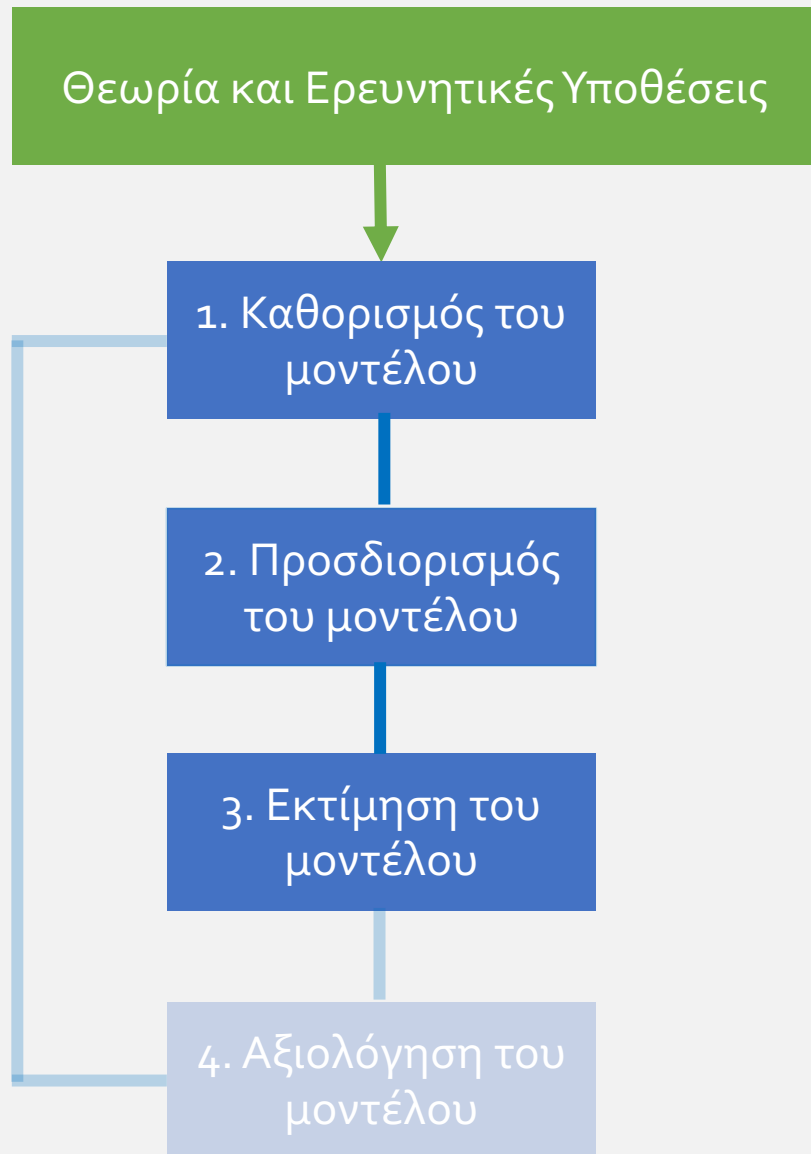
	x1	x2	x3	x4
x1	4.667	1.542	2.458	1.375
x2	1.542	3.523	1.965	.583
x3	2.458	1.965	3.657	1.375
x4	1.375	.583	1.375	2.583

Θεωρητικός πίνακας
διακύμανσης -
συνδιακύμανσης, **Σ**

	x1	x2	x3	x4
x1	4.480	1.493	2.391	1.068
x2	1.493	3.382	1.849	.826
x3	2.391	1.849	3.510	1.323
x4	1.068	.826	1.323	2.480

Διαφορές
(Υπόλοιπα
του μοντέλου)
Residuals

	x1	x2	x3	x4
x1	.187	.049	.067	.307
x2	.049	.141	.116	-.243
x3	.067	.116	.147	.052
x4	.307	-.243	.052	.103



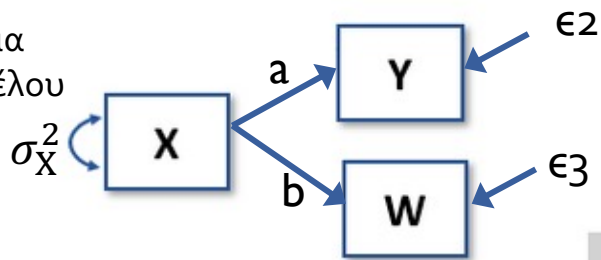
Στόχος της εκτίμησης του μοντέλου είναι η μείωση της απόκλισης του παρατηρούμενου πίνακα διακύμανσης – συνδιακύμανσης και του πίνακα διακύμανσης – συνδιακύμανσης που απορρέει από το μοντέλο και υποθέτουμε ότι ισχύει για τον πληθυσμό.

Μέθοδοι εκτίμησης

- Maximum likelihood (ML): προϋποθέτει συνεχείς μεταβλητές και πολυμεταβλητή Κανονικότητα
- Weighted least squares (WLS)
- Generalized Least Squares (GLS)...

Η εκτίμηση των μοντέλων μέτρησης (CFA) προηγείται της εκτίμησης του δομικού μέρους και γίνεται σε δύο ξεχωριστά βήματα.

διάγραμμα
του μοντέλου



Έστω ότι τα δεδομένα μας γεννήθηκαν από ένα μαθηματικό μοντέλο, το οποίο είναι σε αντιστοιχία με τις ερευνητικές μας υποθέσεις. Το μοντέλο μπορεί να αναπαρασταθεί διαγραμματικά.

Το παραπάνω μοντέλο αντιστοιχεί σε έναν πίνακα **διακυμάνσης – συνδιακυμάνσης** με συγκεκριμένη δομή.

πίνακας
διακυμάνσης -
συνδιακύμανσης
του μοντέλου
Σ

	W	X	Y
W	$b^2\sigma_X^2 + \epsilon_3$		
X	$b^2\sigma_X^2$	σ_X^2	
Y	0	$a^2\sigma_X^2$	$a^2\sigma_X^2 + \epsilon_2$

πίνακας
διακύμανσης –
συνδιακύμανσης
του δείγματος
S

	W	X	Y
W	1.32		
X	0.61	1.40	
Y	0.53	0.74	1.14

*Model estimation tries to match
the data as close as possible*

$$\begin{aligned}\epsilon_2 &= 0.32 \\ \sigma_W^2 &= 1.32 \\ a &= 1.40 \\ b &= 1.14 \\ \epsilon_3 &= 0.51\end{aligned}$$

Οι παράμετροι του μοντέλου εκτιμώνται έτσι ώστε οι δύο πίνακες να έχουν τη μικρότερη δυνατή απόκλιση, δηλαδή .

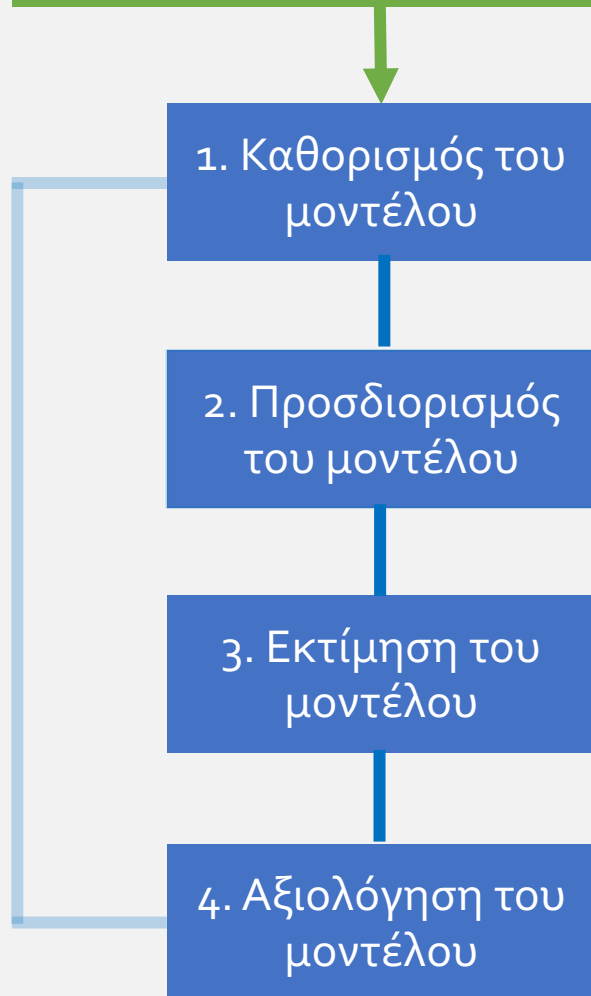
$$\Sigma \approx S$$

διαφορά
(υπόλοιπα)

	W	X	Y
W	0.00		
X	0.61	0.00	
Y	0.53	0.74	0.00

Αξιολογούμε την προσαρμογή του μοντέλου στα δεδομένα. Αν ο **Σ** είναι κοντά στον **S** τότε το μοντέλο προσαρμόζεται καλά στα δεδομένα.

Θεωρία και Ερευνητικές Υποθέσεις



Αξιολόγηση της προσαρμογής του μοντέλου:

- Δείκτες απόλυτης προσαρμογής (στατιστικό χ^2 , GFI, RMSEA, SRMR...)
- Δείκτες σχετικής προσαρμογής (CFI, TLI...)
- Δείκτες φειδωλότητας (AIC, BIC...)

Συνήθως αναφέρουμε: $\chi^2(df)$, p -value (μη στατ. σημαντικό)

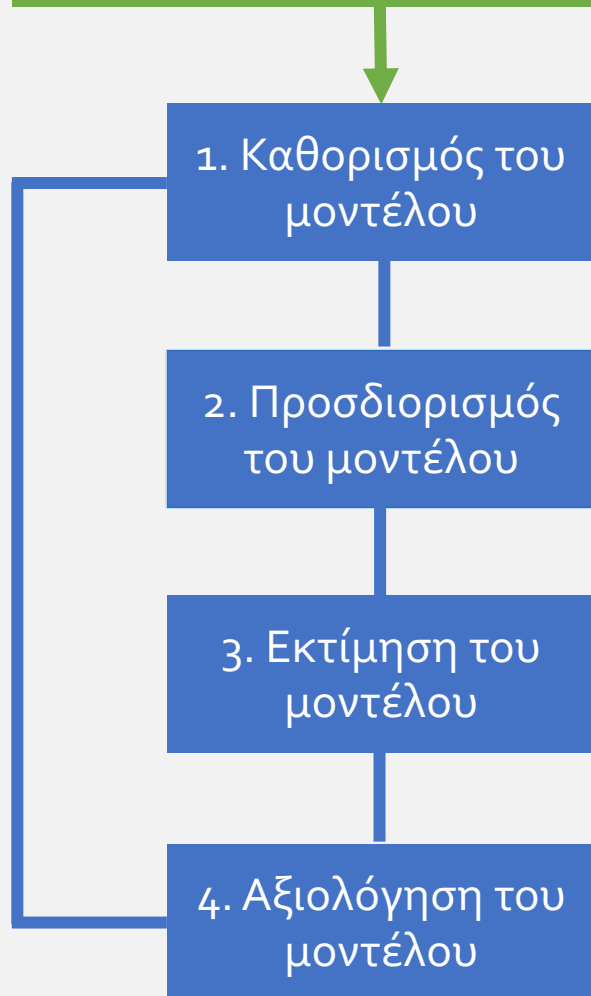
CFI/TLI > .95

RMSEA < .06

SRMR < .08

(Hu & Bentler, 1999)

Θεωρία και Ερευνητικές Υποθέσεις



Τροποποίηση του μοντέλου

Οι όποιες αλλαγές στο μοντέλο πρέπει να έχουν θεωρητικό έρεισμα.

- Δείκτες τροποποίησης (modification indices) του μοντέλου
- Επισκόπηση του πίνακα υπολοίπων του μοντέλου (residuals matrix).

ΠΡΟΣΟΧΗ "The use of SEM has turned into a kind of fetish or numerical trick whose ultimate aim is to maximize the model fit to the data instead of using carefully differentiated research plans and careful substantive considerations on the grounds of theory" (Tarka, 2017)

ΤΙ ΚΑΝΟΥΜΕ ΑΝ...

...δεν ισχύει η προϋπόθεση της Πολυμεταβλητής Κανονικότητας

- Χρήση μεθόδων εκτίμησης που δεν προϋποθέτουν πολυμεταβλητή κανονικότητα
- Χρήση μεθόδων εκτίμησης για κατηγορικά δεδομένα
- Bootstrapping

..έχουμε ελλείπουσες τιμές

- Full-information-maximum-likelihood
- Two-stage
- Multiple imputation

...έχουμε εξαρτημένες παρατηρήσεις (ιεραρχικά δεδομένα)

- Πολυεπίπεδα (Multilevel) SEM

lavaan:

- `estimator="WLS", "WLSMV"`
- `estimator="MLM", "MLR"`
- `se="bootstrap"`

lavaan:

- `missing="FIML"`
- `missing="two.stage"`
- Συμπλήρωση των ελλειπουσών τιμών με χρήση του πακέτου `mice`
συνάρτηση `runMI` του πακέτου `semTools`

lavaan:

- `cluster="..."` + καθορισμός του μοντέλου ανά επίπεδο
- `lavaan.survey`

ΜΕΓΕΘΟΣ ΔΕΙΓΜΑΤΟΣ

Μέθοδος εκτίμησης	Ελάχιστο απαιτούμενο μέγεθος δείγματος
ML με πολυμεταβλητή κανονικότητα	>100 200-400 5:1 λόγος υποκειμένων με αριθμό εκτιμώμενων παραμέτρων 10:1 λόγος υποκειμένων με αριθμό εκτιμώμενων παραμέτρων
MLM για μη κανονικά δεδομένα	>250
Bootstrap με μη κανονικά δεδομένα	> 200 – 1000
Έλεγχοι bootstrap για έμμεσες επιδράσεις	>50 - 500
MLR για μη κανονικά δεδομένα με ελλείπουσες τιμές	> 400
WLSMV	> 200 - 500

https://web.pdx.edu/~newsomj/semclass/ho_sample%20size.pdf

ΛΟΓΙΣΜΙΚΟ ΓΙΑ SEM

✓ εμπορικό – κλειστού κώδικα

- LISREL (Jöreskog & van Thillo, 1972), EQS, AMOS, Mplus (Muthén & Muthén, 2010)
- proc CALIS (SAS/Stat), SEPATH (Statistica), RAMONA (Systat), Stata
- Mx

✓ μη εμπορικό – ανοικτού κώδικα

- πακέτα της R: sem (Fox, 2006), OpenMx (Neale et al., 2016), **lavaan** (Rosseel, 2012), lava, psychonetrics
- gllamm (Stata), Onyx, semopy (Igolkina & Meshcheryakov, 2020), κ.ά.



- Η R είναι **ανοικτό λογισμικό** και **δωρεάν στη χρήση** της από οποιονδήποτε (ιδιώτη ή επιχείρηση). Είναι διαθέσιμη για όλα τα υπολογιστικά περιβάλλοντα και λειτουργικά συστήματα.
- Έχει σχεδιαστεί με στόχο την υποστήριξη επεξεργασίας, εξερεύνησης και ανάλυσης δεδομένων.
- Αποτελεί μία ολοκληρωμένη και πλήρης στατιστική πλατφόρμα υποστηρίζοντας όλες τις τεχνικές και μεθόδους ανάλυσης δεδομένων. Οποιοσδήποτε μπορεί να αναπτύξει νέα **βιβλιοθήκη**, η οποία περιέχει **συναρτήσεις** για την **εκτέλεση οποιασδήποτε στατιστικής μεθόδου**.
- Μπορεί με ευκολία να ενσωματώσει και να χρησιμοποιήσει δεδομένα που υπάρχουν σε εξωτερικές (τρίτες) πηγές που μπορούν να είναι σε οποιαδήποτε μορφή (όπως αρχεία κειμένου, αρχεία Excel, βάσεις δεδομένων κλπ).

LAVAAN - LATENT VARIABLE ANALYSIS

<https://lavaan.ugent.be/>

Το **lavaan** είναι ένα πακέτο της R για την ανάλυση λανθανουσών μεταβλητών:

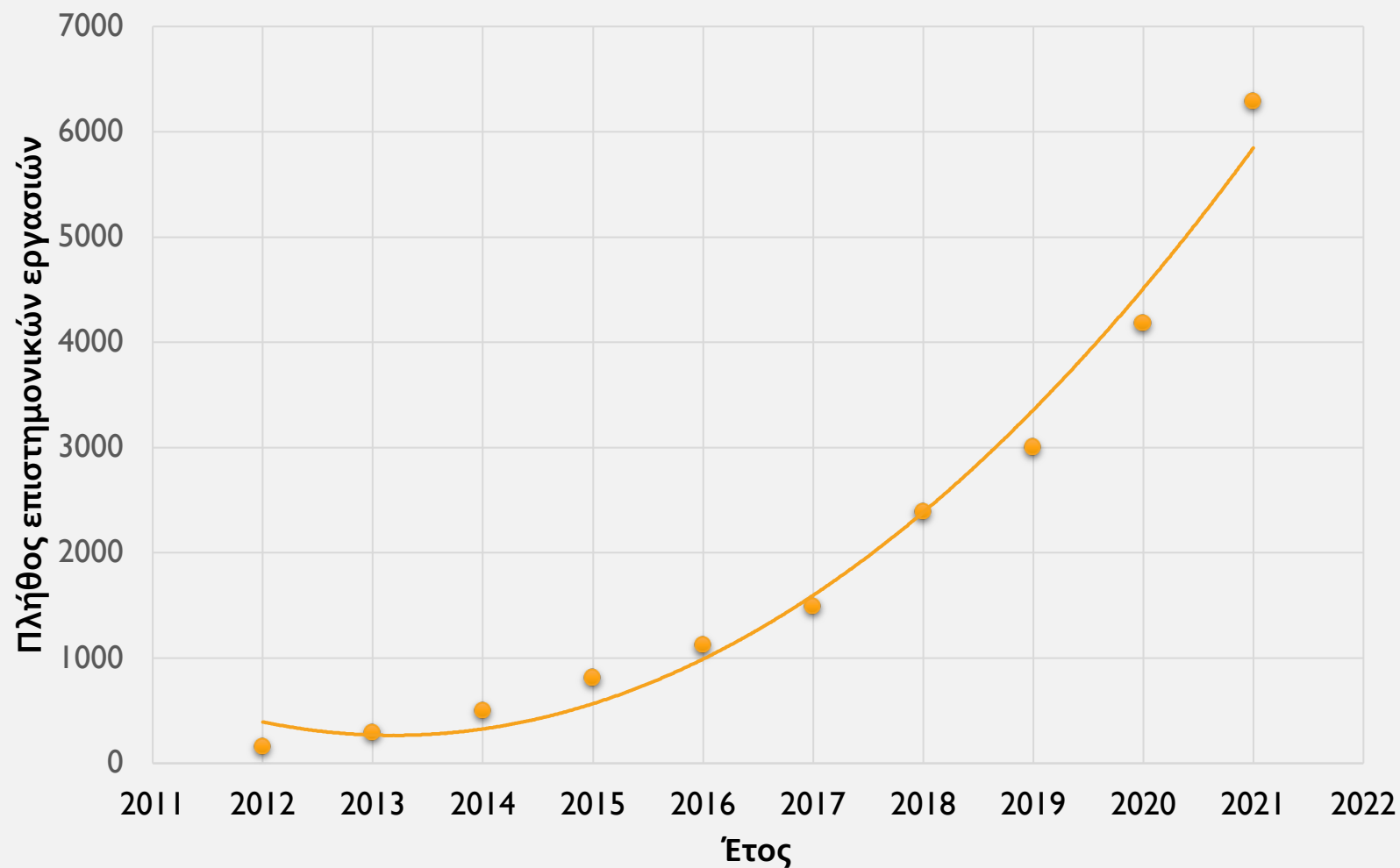
πλεονεκτήματα:

- συντακτικό φιλικό στο χρήστη
- διαχείριση ελλειπουσών τιμών
- συνεχείς, διχοτομικές και τακτικές μεταβλητές στο μοντέλο
- μη κανονικές κατανομές, ιεραρχικά δεδομένα

μέθοδοι:

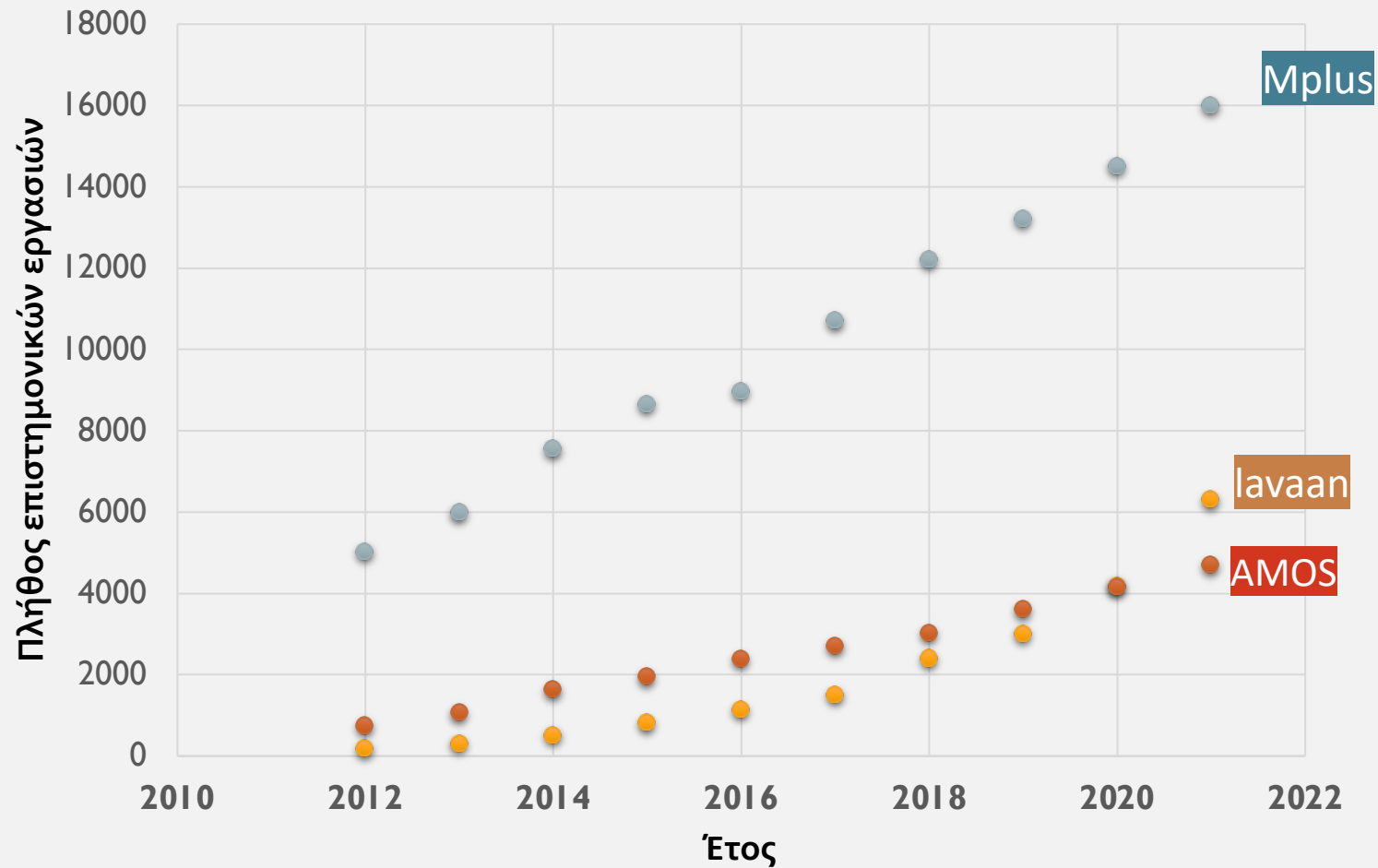
- path analysis, CFA, EFA, multiple group CFA, SEM, multilevel SEM, ESEM, growth curve models, IRT

ΔΗΜΟΤΙΚΟΤΗΤΑ



Πλήθος επιστημονικών εργασιών που αναφέρουν το lavaan (2012 – 2021)
(Πηγή: Google Scholar)

LAVAAN VS AMOS VS MPLUS



Πλήθος επιστημονικών εργασιών που αναφέρουν τα Mplus, lavaan και SPSS AMOS (2012 – 2021). Πηγή: Google Scholar

ΒΙΒΛΙΟΓΡΑΦΙΑ

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., ... & Moffitt, T. E. (2014). The p factor: one general psychopathology factor in the structure of psychiatric disorders?. *Clinical Psychological Science*, 2(2), 119-137.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: a Multidisciplinary Journal*, 6(1), 1-55.

Kline, R.B. (2021). *Μοντέλα Δομικών Εξισώσεων*. Η. Σαντουρίδης, & Π. Πολυχρονίδου (Επιμ.). Αθήνα: Προπομπός.

Mair, P. (2018). *Modern Psychometrics with R*. Cham: Springer International Publishing.

Rindermann, H., & Neubauer, A. C. (2004). Processing speed, intelligence, creativity, and school performance: Testing of causal hypotheses using structural equation models. *Intelligence*, 32(6), 573-589.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling, *Journal of Statistical Software*, 48(2), 1-36.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Tarka, P. (2018). An overview of structural equation modeling: its beginnings, historical development, usefulness and controversies in the social sciences, *Quality & Quantity*, 52(1), 313–354.

Willis, M., & Jozkowski, K. N. (2022). Sexual consent perceptions of a fictional vignette: A latent growth curve model. *Archives of Sexual Behavior*, 51(2), 797-809.

Zhang, J., Cao, C., Shen, S., & Qian, M. (2019). Examining effects of self-efficacy on research motivation among chinese university teachers: Moderation of leader support and mediation of goal orientations. *The Journal of Psychology*, 153(4), 414-435

Ζαφειρόπουλος, Κ. (2012). *Ποσοτική εμπειρική έρευνα και δημιουργία στατιστικών μοντέλων*. Αθήνα: Κριτική.

Τσιγγίλης, Ν. (2010). Βασικές έννοιες και εφαρμογή της μοντελοποίησης δομικών εξισώσεων. Στο Π. Μεταλλίδου, Π. Ρούσση, Α. Μπρούζος & Α. Ευκλείδη (Επ. Έκδ.), *Επιστημονική Επετηρίδα Ψυχολογικής Εταιρείας Βορείου Ελλάδας*, 8, 37-67.

ΕΥΧΑΡΙΣΤΩ ΓΙΑ ΤΗΝ ΠΡΟΣΟΧΗ ΣΑΣ.
ΕΡΩΤΗΣΕΙΣ;



Credit: DALL-E



<https://github.com/amarkos/semworkshop>

Πρόσβαση στο υλικό του σεμιναρίου