

Άγγελος Μάρκος amarkos.gr

ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΔΟΜΙΚΩΝ ΕΞΙΣΩΣΕΩΝ

ME TH XPHΣH THΣ R (EPΓΑΣΤΗΡΙΟ)





https://github.com/amarkos/semworkshop

Πρόσβαση στο υλικό του σεμιναρίου

ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΔΟΜΙΚΩΝ ΕΞΙΣΩΣΕΩΝ

ME TH XPHΣHTHΣ R
(ΕΡΓΑΣΤΗΡΙΟ)

ΠΕΡΙΕΧΟΜΕΝΑ

- Μια (πολύ) σύντομη εισαγωγή στην R
- Εφαρμογή σε πραγματικά δεδομένα με το lavaan

ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΔΟΜΙΚΩΝ ΕΞΙΣΩΣΕΩΝ

ME TH XPHΣH THΣ R (EPΓΑΣΤΗΡΙΟ)

ΜΙΑ (ΠΟΛΥ) ΣΥΝΤΟΜΗ ΕΙΣΑΓΩΓΗ ΣΤΗΝ R

- Οδηγίες εγκατάστασης της R και του Rstudio
 https://static.eudoxus.gr/books/https://static.eudoxus.gr/books/92/ch
 apter-86197192.pdf
- ο Εισαγωγή δεδομένων (SEM.R)
- ο Βασικές δομές δεδομένων (αρχείο R-Data-Structures.pdf)

ΤΑ ΔΕΔΟΜΕΝΑ

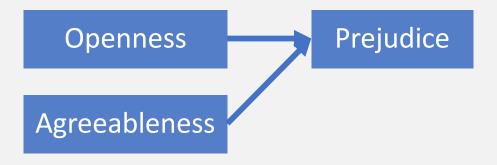
Journal of Personality and Social Psychology 2016, Vol. 111, No. 3, 367-395 © 2016 American Psychological Association 0022-3514/16/\$12.00 http://dx.doi.org/10.1037/pspi0000064

Is Group Membership Necessary for Understanding Generalized Prejudice?

A Re-Evaluation of Why Prejudices Are Interrelated

Robin Bergh Harvard University and Uppsala University Nazar Akrami Uppsala University

Jim Sidanius Harvard University Chris G. Sibley University of Auckland



R package MPsychoR data (Bergh)

Format

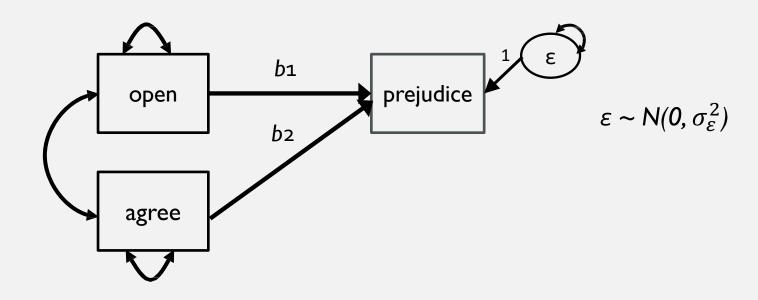
A data frame with 861 individuals, 10 composite scores, and gender:

- EP Ethnic prejudice
- SP Sexism
- HP Sexual prejudice against gays and lesbians
- DP Prejudice toward mentally people with disabilities
- A1 Agreeableness indicator 1
- A2 Agreeableness indicator 2
- A3 Agreeableness indicator 3
- 01 Openness indicator 1
- 02 Openness indicator 2
- 03 Openness indicator 3

gender gender

MODEL 1

Μοντέλο γραμμικής παλινδρόμησης με δύο ανεξάρτητες μεταβλητές



Prejudice (Προκατάληψη) Openness to experience (Ανοικτότητα στην εμπειρία) Agreeableness (Προσήνεια)

Θεωρία και Ερευνητικές Υποθέσεις 1. Καθορισμός του μοντέλου

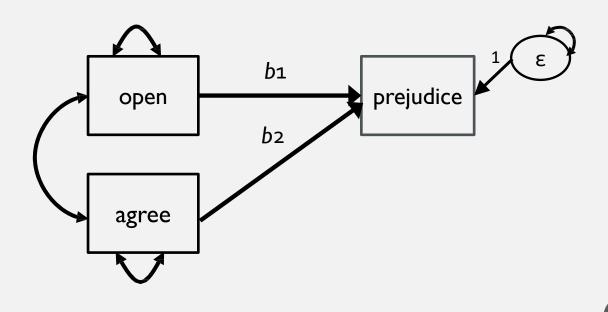
lavaan:

- Παλινδρόμηση: Υ ~ Χ
- Συνδιακύμανση: Υ ~ ~ Χ
- Λανθάνουσα μεταβλητή: eta =~ x1 + x2 + x3

```
#Φόρτωση δεδομένων
data(Bergh)
View(Bergh)
attach(Bergh)
```

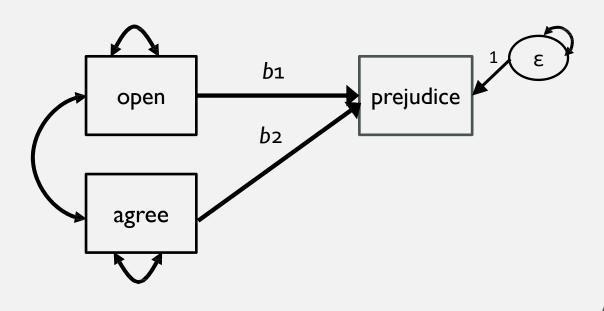
Θεωρία και Ερευνητικές Υποθέσεις 1. Καθορισμός του μοντέλου

Δημιουργία σύνθετων μεταβλητών
Bergh\$Open <- (01+02+03)/3
Bergh\$Agree <- (A1+A2+A3)/3
Bergh\$Prejudice <- (EP+SP+DP+HP)/4



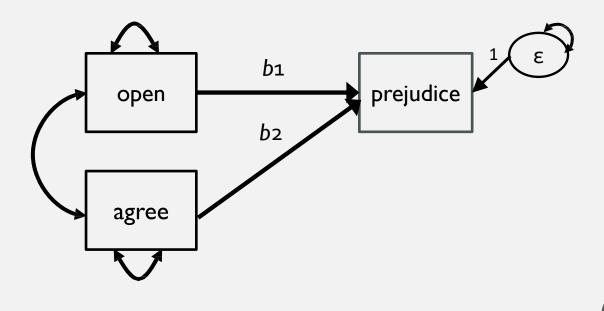
Θεωρία και Ερευνητικές Υποθέσεις 1. Καθορισμός του μοντέλου

```
# Βήμα 1: Καθορισμός του μοντέλου
model1 <- '
# Δομικό μοντέλο
Prejudice ~ b1*Open + b2*Agree
# Διακυμάνσεις/Συνδιακυμάνσεις
Open ~~ Open + Agree
Agree ~~ Agree'
```



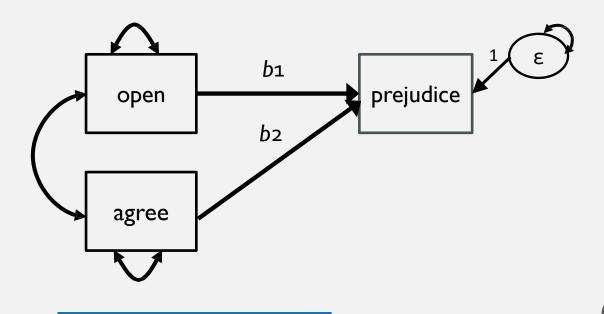
Θεωρία και Ερευνητικές Υποθέσεις 1. Καθορισμός του μοντέλου 2. Προσδιορισμός του μοντέλου 3. Εκτίμηση του μοντέλου

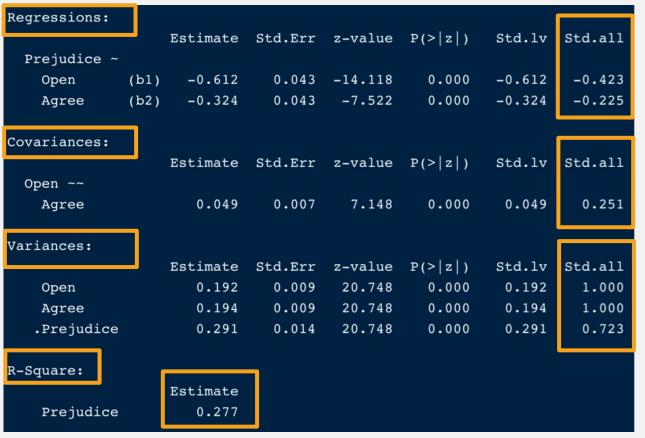
```
# Βήμα 2: Εκτίμηση του μοντέλου model1.fit <- sem(model1, data = Bergh, meanstructure = FALSE, estimator = "ML")
```

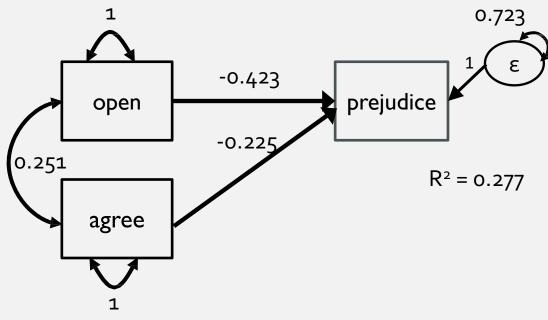


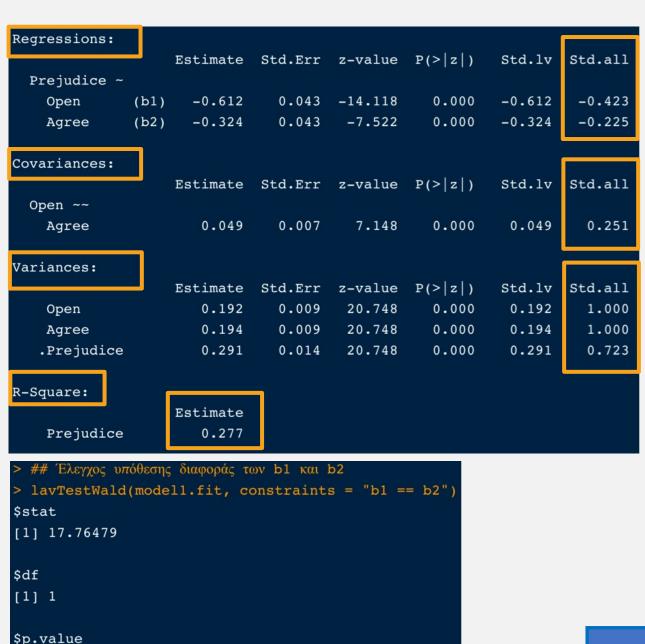
Θεωρία και Ερευνητικές Υποθέσεις 1. Καθορισμός του μοντέλου 2. Προσδιορισμός του μοντέλου 3. Εκτίμηση του μοντέλου 4. Αξιολόγηση του μοντέλου

Βήμα 3: Αξιολόγηση του μοντέλου summary(model1.fit, rsquare = TRUE, fit.measures = TRUE, standardized = TRUE)

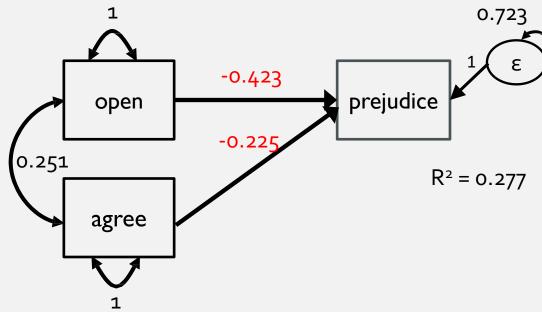






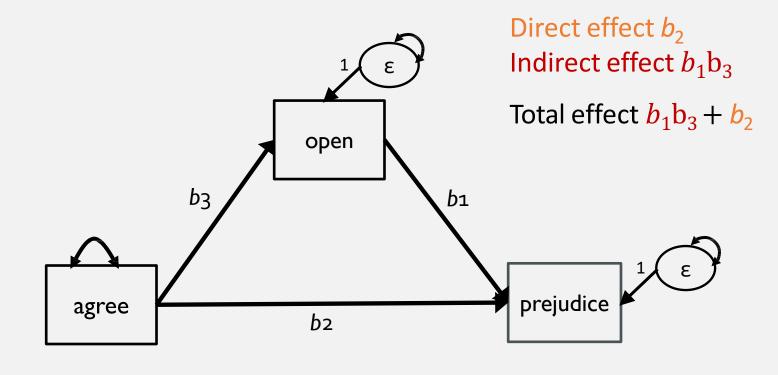


[1] 2.499661e-05



MODEL 2

Μοντέλο διαμεσολάβησης με άμεσες και έμμεσες επιδράσεις

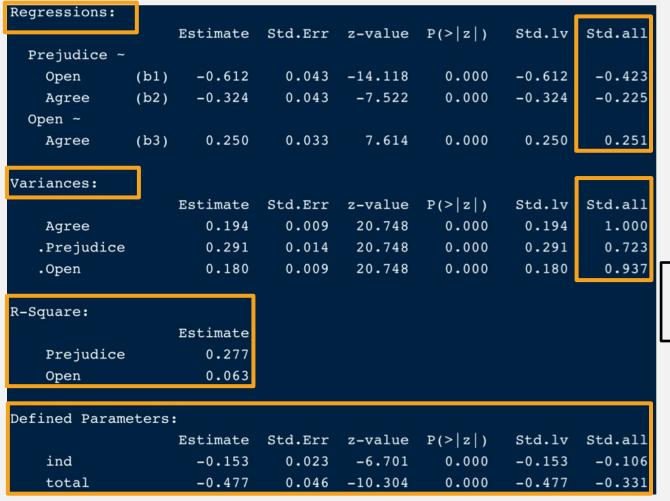


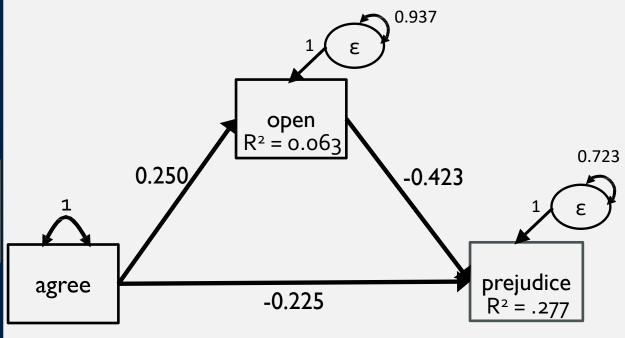
Prejudice (Προκατάληψη) Openness to experience (Ανοικτότητα στην εμπειρία) Agreeableness (Προσήνεια)

MODEL 2

Μοντέλο διαμεσολάβησης με άμεσες και έμμεσες επιδράσεις

```
model2 <- '
# Structural model
Prejudice ~ b1*Open + b2*Agree
Open ~ b3*Agree
# New parameters
# indirect effect
ind := b1*b3
# total effect
total := b2 + (b1*b3)
# Βήμα 2: Εκτίμηση του μοντέλου
model2.fit <- sem(model2,
        data = Bergh,
        estimator = "ML")
# Βήμα 3: Αξιολόγηση του μοντέλου
summary(model2.fit,
   rsquare = TRUE,
   fit.measures = TRUE,
   standardized = TRUE)
```





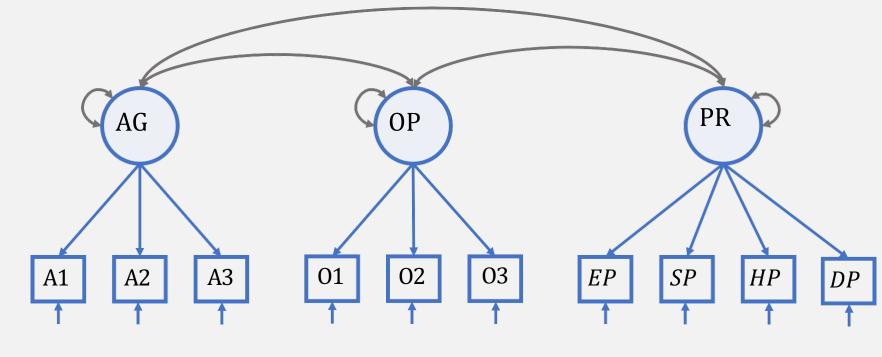
Indirect effect $b_1b_3 = -0.106$

Total effect =
$$b_2 + (b_1b_3) = -0.331$$

MODEL 3A

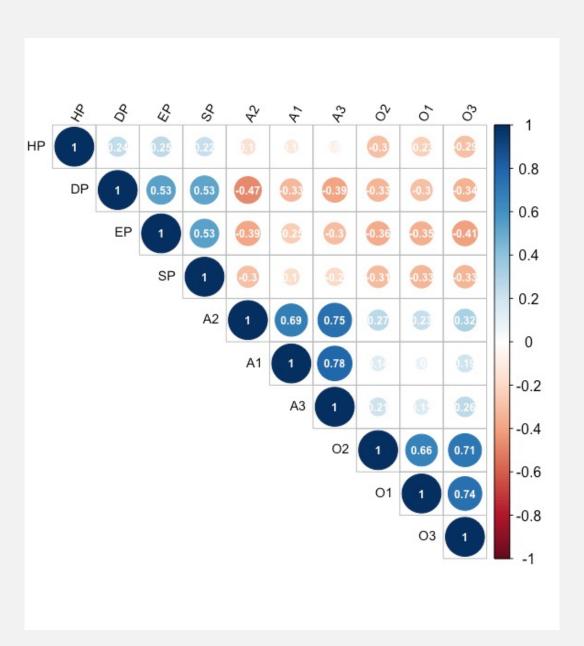
Επιβεβαιωτική ανάλυση παραγόντων (CFA) με τρεις συσχετισμένες λανθάνουσες μεταβλητές ή παράγοντες

Αναπαριστούμε τις εννοιολογικές κατασκευές ως λανθάνουσες μεταβλητές και διαχωρίζουμε την «πραγματική τιμή» από το σφάλμα της μέτρησης.



MODEL 3A

Συσχετίσεις



MODEL 3A

Επιβεβαιωτική ανάλυση παραγόντων (CFA) με τρεις συσχετισμένες λανθάνουσες μεταβλητές ή παράγοντες

```
# Βήμα 1: Καθορισμός του μοντέλου
model3 <- '
# Measurement models
OP = ~O1 + O2 + O3
AG = -A_1 + A_2 + A_3
PR = EP + SP + HP + DP
# Covariance structure
OP ~~ OP + AG + PR
AG ~~ AG + PR'
# Βήμα 2: Εκτίμηση του μοντέλου
model3.fit <- sem(model3,
data = Bergh,
estimator = "ML")
# Βήμα 3: Αξιολόγηση του μοντέλου
summary(model3.fit,
fit.measures = TRUE,
standardized = TRUE)
```

lavaan 0.6-12 ended normally after 54	iterations
Estimator	ML
Optimization method	NLMINB
Number of model parameters	23
Number of observations	861
Model Test User Model:	
Test statistic	186.620
Degrees of freedom	32
P-value (Chi-square)	0.000
Model Test Baseline Model:	
Test statistic	4270.205
Degrees of freedom	45
P-value	0.000
User Model versus Baseline Model:	
Comparative Fit Index (CFI)	0.963
Tucker-Lewis Index (TLI)	0.949

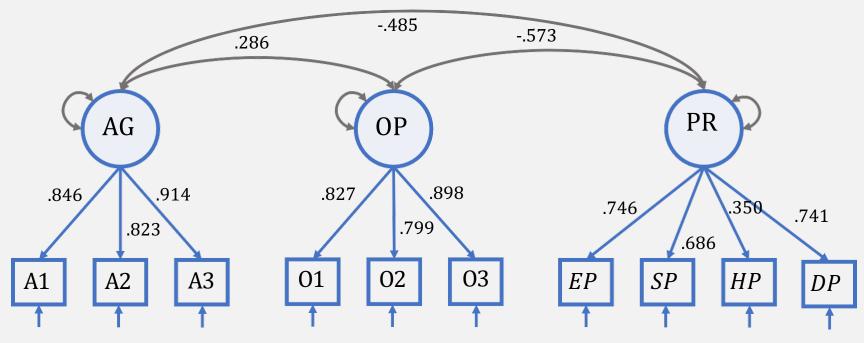
Loglikelihood and Information Criteria:	
Loglikelihood user model (H0)	-5672.807
Loglikelihood unrestricted model (H1)	-5579.497
Akaike (AIC)	11391.614
Bayesian (BIC)	11501.050
Sample-size adjusted Bayesian (BIC)	11428.008
Root Mean Square Error of Approximation:	
Root Mean Square Error of Approximation:	
Root Mean Square Error of Approximation: RMSEA 90 Percent confidence interval - lower	0.075 0.065
RMSEA	
RMSEA 90 Percent confidence interval - lower	0.065
RMSEA 90 Percent confidence interval - lower 90 Percent confidence interval - upper	0.065 0.085

 $\chi^2(32) = 186.62$, CFI = .96, TLI = .95, RMSEA = .075 (.065 - .085), SRMR = .054

Το μοντέλο προσαρμόζεται ικανοποιητικά στα δεδομένα. Έχουμε ισχυρές ενδείξεις ότι υποστηρίζεται η παραγοντική δομή που υποθέσαμε.

Latent Variables:						
	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
OP =~						
01	1.000				0.400	0.827
02	0.934	0.036	26.185	0.000	0.374	0.799
03	1.149	0.040	28.900	0.000	0.460	0.898
AG =~						
A1	1.000				0.426	0.846
A2	0.910	0.032	28.812	0.000	0.388	0.823
A3	1.030	0.032	31.899	0.000	0.439	0.914
PR =~						
EP	1.000				0.530	0.746
SP	0.886	0.051	17.348	0.000	0.469	0.686
HP	1.030	0.112	9.160	0.000	0.545	0.350
DP	0.746	0.041	18.308	0.000	0.395	0.741
Covariances:						
	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
OP ~~						
AG	0.049	0.007	7.105	0.000	0.286	0.286
PR	-0.122	0.011	-11.371	0.000	-0.573	-0.573
AG ~~						
PR	-0.110	0.011	-10.241	0.000	-0.485	-0.485

Τα παραγοντικά φορτία είναι υψηλά (>0.3) και κυμαίνονται από 0.350 έως 0.898.



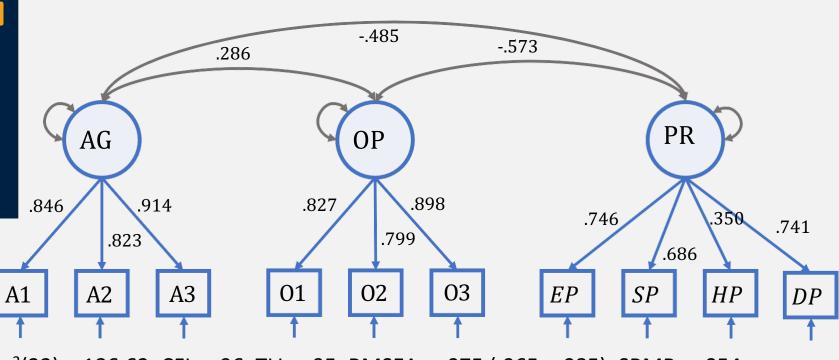
 $\chi^{2}(32) = 186.62$, CFI = .96, TLI = .95, RMSEA = .075 (.065 - .085), SRMR = .054

Δείκτες τροποποίησης του μοντέλου (modification indices):

> 1	modir	ndio	ces(r	nodel3a	fit,son	rt = TRUE	E, minimur	m.value =	10
	lhs	ор	rhs	mi	ерс	sepc.lv	sepc.all	sepc.nox	
72	A1	~~	A3	69.141	0.088	0.088	1.688	1.688	
45	PR	=~	A2	68.775	-0.220	-0.117	-0.247	-0.247	
28	OP	=~	A2	34.333	0.169	0.068	0.144	0.144	
40	AG	=~	DP	32.769	-0.268	-0.114	-0.214	-0.214	
38	AG	=~	SP	22.259	0.283	0.121	0.176	0.176	
32	OP	=~	HP	22.203	-0.842	-0.337	-0.216	-0.216	
77	A2	~~	A3	19.619	-0.041	-0.041	-0.778	-0.778	
44	PR	=~	A1	19.206	0.121	0.064	0.127	0.127	
27	OP	=~	A1	19.169	-0.131	-0.052	-0.104	-0.104	
34	AG	=~	01	15.221	-0.111	-0.047	-0.098	-0.098	
36	AG	=~	03	10.358	0.094	0.040	0.078	0.078	
				<u> </u>	<u>"</u>		·		

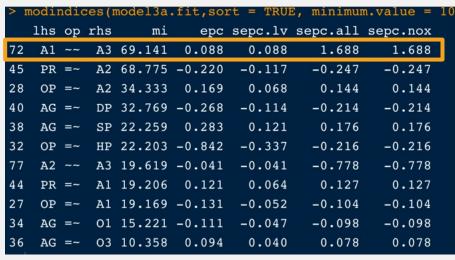
Η τιμή ενός δείκτη τροποποίησης (στήλη mi) μας λέει πόσο θα μειωθεί το στατιστικό χ² αν ελευθερώσουμε την αντίστοιχη παράμετρο.

Το νέο μοντέλο επομένως, αναμένεται να προσαρμόζεται καλύτερα στα δεδομένα.



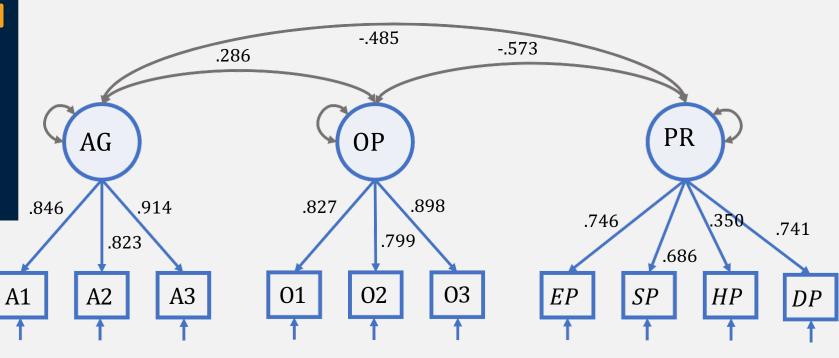
 $\chi^2(32)$ = 186.62, CFI = .96, TLI = .95, RMSEA = .075 (.065 - .085), SRMR = .054

Δείκτες τροποποίησης του μοντέλου (modification indices):



Μπορεί να τεκμηριωθεί θεωρητικά αυτή η τροποποίηση;

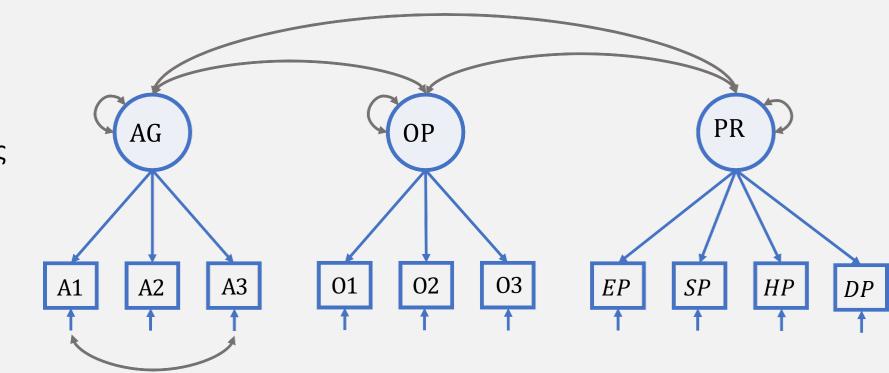
Τα σφάλματα υπολοίπων για τους δείκτες Α1 και Α3 συσχετίζονται μεταξύ τους λόγω της παρόμοιας διατύπωσης των σχετικών ερωτήσεων (item wording)!



 $\chi^{2}(32) = 186.62$, CFI = .96, TLI = .95, RMSEA = .075 (.065 - .085), SRMR = .054

MODEL 3B

Επιβεβαιωτική ανάλυση παραγόντων (CFA) με τρεις συσχετισμένες λανθάνουσες μεταβλητές ή παράγοντες (τροποποιημένο μοντέλο)



Residual covariance AI ~~ A3

lavaan 0.6-12 ended normally after 62	iterations
Estimator	ML
Optimization method	NLMINB
Number of model parameters	24
Number of observations	861
Model Test User Model:	
Test statistic	118.256
Degrees of freedom	31
P-value (Chi-square)	0.000
Model Test Baseline Model:	
Test statistic	4270.205
Degrees of freedom	45
P-value	0.000
User Model versus Baseline Model:	
Comparative Fit Index (CFI)	0.979
Tucker-Lewis Index (TLI)	0.970

Loglikelihood and Information Criteria:	
Loglikelihood user model (H0)	-5638.625
Loglikelihood unrestricted model (H1)	-5579.497
Akaike (AIC)	11325.249
Bayesian (BIC)	11439.444
Sample-size adjusted Bayesian (BIC)	11363.226
RMSEA 90 Percent confidence interval - lower 90 Percent confidence interval - upper P-value RMSEA <= 0.05	0.057 0.046 0.068 0.131
1-varae mibbh <- 0:03	0.131

 $\chi^2(31)$ = 118.26, CFI = .98, TLI = .97, RMSEA = .057 (.046 - .068), SRMR = .043 Το μοντέλο προσαρμόζεται καλά στα δεδομένα.

Latent Variables:						
A PANIS COMMING ON THE PANIS CONTRACT OF THE	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
OP =~						
01	1.000				0.400	0.827
02	0.934	0.036	26.188	0.000	0.374	0.799
03	1.149	0.040	28.921	0.000	0.460	0.898
AG =~						
A1	1.000				0.346	0.687
A2	1.361	0.086	15.756	0.000	0.471	0.999
A3	1.036	0.033	31.662	0.000	0.358	0.746
PR =~						
EP	1.000				0.529	0.744
SP	0.887	0.051	17.460	0.000	0.469	0.685
HP	1.031	0.112	9.177	0.000	0.545	0.349
DP	0.750	0.040	18.535	0.000	0.397	0.744
Covariances:						
	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
OP ~~						
AG	0.046	0.006	7.516	0.000	0.330	0.330
PR	-0.121	0.011	-11.372	0.000	-0.573	-0.573
AG ~~						
PR	-0.098	0.010	-9.409	0.000	-0.536	-0.536
.A1 ~~						
.A3	0.066	0.008	8.266	0.000	0.066	0.560

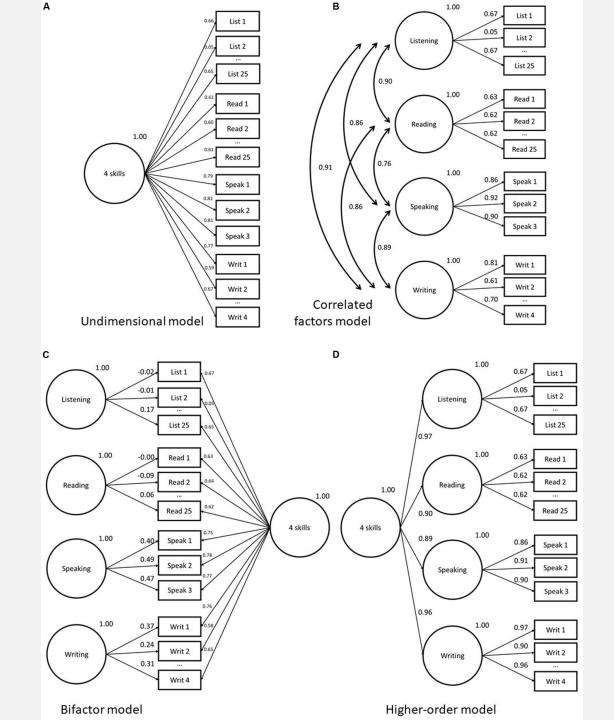
Είναι όμως η προσαρμογή του μοντέλου 3B σημαντικά καλύτερη από αυτή του 3A;

Τα δύο μοντέλα είναι εμφωλευμένα (nested): το 3Β είναι ίδιο με το 3Α, με έναν επιπλέον περιορισμό.

Ναι.

```
> ## Σύγκριση των μοντέλων: Model 3 vs. refined Model 3
> anova(model3.fit, model3b.fit)
Chi-Squared Difference Test
                  AIC BIC Chisq Chisq diff Df diff Pr(>Chisq)
model3b.fit 31 11325 11439 118.26
model3.fit 32 11392 11501 186.62 68.364 1 < 2.2e-16 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
MODEL 3B: \chi^2(31) = 118.26, CFI = .98, TLI = .97, RMSEA = .057 (.046 - .068), SRMR = .043
MODEL 3A: \chi^2(32) = 186.62, CFI = .96, TLI = .95, RMSEA = .075 (.065 - .085), SRMR = .054
\Delta \chi^2(1) = 68.36, p < 0.001
```

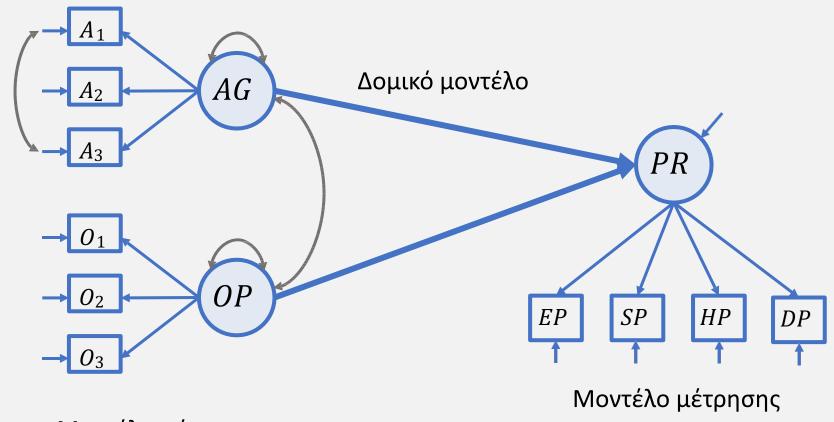
ΑΛΛΑ ΜΟΝΤΕΛΑ ΜΕΤΡΗΣΗΣ



MODEL 4

Moντέλο SEM των Bergh et al. (2016)

Το μοντέλο αυτό περιέχει και δομικό μέρος και μέρος που αφορά στη μέτρηση.



Μοντέλα μέτρησης

MODEL 4

Μοντέλο δομικών εξισώσεων με τρεις λανθάνουσες μεταβλητές

```
model4 <- '# Measurement models
    OP = ~O1 + O2 + O3
    AG = -A_1 + A_2 + A_3
    PR = EP + SP + HP + DP
    # Residual covariance
    A1 ~~ A3
    # Structural model
    PR ~ b1*OP + b2*AG'
model4.fit <- sem(model4, data = Bergh, estimator = "ML")
summary(model4.fit, standardized = TRUE, fit.measures =
TRUE, rsquare = TRUE)
```

lavaan 0.6-12 ended normally after 55	iterations					
Estimator	ML					
Optimization method	NLMINB					
Number of model parameters	24					
Number of observations	861					
Model Test User Model:						
Test statistic	118.256					
Degrees of freedom	31					
P-value (Chi-square)	0.000					
Model Test Baseline Model:						
Test statistic	4270.205					
Degrees of freedom	45					
P-value	0.000					
User Model versus Baseline Model:						
Comparative Fit Index (CFI)	0.979					
Tucker-Lewis Index (TLI)	0.970					

Loglikelihood and Information Criteria:	
Loglikelihood user model (H0)	-5638.625
Loglikelihood unrestricted model (H1)	-5579.497
Akaike (AIC)	11325.249
Bayesian (BIC)	11439.444
Sample-size adjusted Bayesian (BIC)	11363.226
RMSEA 90 Percent confidence interval - lower	0.057 0.046
90 Percent confidence interval - upper	0.068
P-value RMSEA <= 0.05	0.131
Standardized Root Mean Square Residual:	
SRMR	0.043

 $\chi^2(31) = 118.256$, CFI = .98, TLI = .97, RMSEA = .057 (.046 - .068), SRMR = .043

Το μοντέλο προσαρμόζεται ικανοποιητικά στα δεδομένα. Έχουμε ισχυρές ενδείξεις ότι υποστηρίζεται η παραγοντική δομή που υποθέσαμε.

Latent Variab	les:						
		Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
OP =~							
01		1.000				0.400	0.827
02		0.934	0.036	26.188	0.000	0.374	0.799
03		1.149	0.040	28.921	0.000	0.460	0.898
AG =~							
A1		1.000				0.346	0.687
A2		1.361	0.086	15.756	0.000	0.471	0.999
A3		1.036	0.033	31.662	0.000	0.358	0.746
PR =~							
EP		1.000				0.529	0.744
SP		0.887	0.051	17.460	0.000	0.469	0.685
HP		1.031	0.112	9.177	0.000	0.545	0.349
DP		0.750	0.040	18.535	0.000	0.397	0.744
Regressions:							
		Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
PR ~							
OP	(b1)	-0.587	0.053	-11.106	0.000	-0.444	-0.444
AG	(b2)	-0.595	0.058	-10.172	0.000	-0.390	-0.390
Covariances:							
		Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
.A1 ~~							
.A3 OP ~~		0.066	0.008	8.266	0.000	0.066	0.560
AG		0.046	0.006	7.516	0.000	0.330	0.330

Όταν δεν ισχύει η προϋπόθεση της Κανονικότητας

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

Δεδομένα με ελλείπουσες τιμές

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

Εξαρτημένες παρατηρήσεις (ιεραρχικά δεδομένα)

■ Πολυεπίπεδα (Multilevel) SEM

lavaan:

- estimator="WLS", "WLSMV"
- estimator="MLM", "MLR»
- se="bootstrap"; test="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

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;







https://github.com/amarkos/semworkshop Πρόσβαση στο υλικό του σεμιναρίου