

# ΠΑΡΑΓΟΝΤΙΚΗ ΑΝΑΛΥΣΗ FACTOR ANALYSIS

- ΣΤΟΧΟΣ
- ΜΕΙΟΝΕΚΤΗΜΑΤΑ-ΠΛΕΟΝΕΚΤΗΜΑΤΑ
- ΔΙΑΦΟΡΕΣ ΜΕ ΚΥΡΙΕΣ ΣΥΝΙΣΤΩΣΕΣ.
- ΚΑΤΑΛΛΗΛΟΤΗΤΑ ΔΕΔΟΜΕΝΩΝ.
- ΠΩΣ ΕΠΙΤΥΓΧΑΝΕΤΑΙ.
- ΠΑΡΑΔΕΙΓΜΑ
- ΥΛΟΠΟΙΗΣΗ ΣΤΟ SPSS
- ΕΡΜΗΝΕΙΑ

# ΠΑΡΑΓΟΝΤΙΚΗ ΑΝΑΛΥΣΗ

- **ΣΤΟΧΟΣ:** η εύρεση της ύπαρξης κοινών παραγόντων ανάμεσα σε μία ομάδα μεταβλητών. Τι επιτυγχάνεται?
  1. Μείωση της διάστασης του προβλήματος.
  2. Εξήγηση των συσχετίσεων που υπάρχουν στα δεδομένα.
  3. Δημιουργία νέων μεταβλητών που ίσως ερμηνεύουν μη μετρήσιμες έννοιες!!!

# ΠΑΡΑΓΟΝΤΙΚΗ ΑΝΑΛΥΣΗ

4. Δημιουργία ενός συνόλου παραγόντων για να χρησιμοποιηθούν ως ασυσχέτιστες μεταβλητές (διόρθωση προβλήματος πολυσυγγραμμικότητας)
5. Εξακρίβωση ή επιβεβαίωση μιας κλίμακας.

# ΜΕΙΟΝΕΚΤΗΜΑΤΑ ΠΑΡΑΓΟΝΤΙΚΗΣ

1. Υποθέσεις μη ρεαλιστικές και ελέγξιμες.
2. Δεν δίνει μοναδική λύση.
3. Οι παράγοντες που προκύπτουν δεν έχουν μοναδική ερμηνεία.
4. Ο αριθμός των παραγόντων δεν είναι αυστηρά – μαθηματικά καθορισμένος.

# ΠΑΡΑΓΟΝΤΙΚΗ-ΚΥΡΙΕΣ ΣΥΝΙΣΤΩΣΕΣ

1. Η Π.Α. είναι στατιστική τεχνική ενώ οι Κ.Σ. μαθηματικός μετασχηματισμός.
2. Οι Κ.Σ. εξηγούν την διακύμανση ενώ η Π.Α. την συνδιακύμανση.
3. Στην Π.Α. οι μεταβλητές γράφονται ως γραμμικός συνδυασμός των παραγόντων, ενώ οι Κ.Σ. ως γραμμικός συνδυασμός των μεταβλητών.

# Είδη παραγοντικής ανάλυσης

- Exploratory  
Προσπαθεί να ανακαλύψει την ύπαρξη παραγόντων σε ένα μεγάλο σύνολο μεταβλητών.
- Confirmatory  
Προσπαθεί να ανακαλύψει αν ο αριθμός των παραγόντων καθώς και η σύνθεση τους επιβεβαιώνει τα αναμενόμενα από τη θεωρία.

# Μαθηματικό μοντέλο (εν συντομία)

$p$  μεταβλητές και  $k$  παράγοντες.

$$X - \mu = LF + \varepsilon$$

$$L_{p \times k} = (L_{ij})$$

Επιβάρυνση του  $j$  παράγοντα στην  $i$  μεταβλητή

$$X_1 = L_{11}F_1 + L_{12}F_2 + \dots + L_{1k}F_k$$

$$X_2 = L_{21}F_1 + L_{22}F_2 + \dots + L_{2k}F_k$$

$$X_p = L_{p1}F_1 + L_{p2}F_2 + \dots + L_{pk}F_k$$

# Υποθέσεις

1.  $E(F)=0$
2.  $\text{Cov}(F)=I$
3.  $E(\varepsilon)=0$
4.  $\text{Cov}(\varepsilon)=\Psi=\text{diag}(\Psi_1, \dots, \Psi_p)$
5. Τα διαγώνια στοιχεία του  $\text{Cov}(e, F)=0$ .
6. Δεδομένα από πολυμεταβλητούς κανονικούς πληθυσμούς (μπορεί να αγνοείται).
7. Όχι ακραίες τιμές
8. Κατάλληλο μέγεθος δείγματος:  $10 \cdot p$  ή  $300+$



# Exploratory factor analysis

1. Καταλληλότητα των δεδομένων
2. Καταλληλότητα των μεταβλητών
3. Αριθμός παραγόντων.
4. Εκτίμηση των παραγόντων.
5. Περιστροφή.
6. Ερμηνεία των παραγόντων

# Καταλληλότητα δεδομένων- μεταβλητών

- Θα πρέπει να υπάρχουν ικανοποιητικές συσχετίσεις για τη διεξαγωγή παραγοντικής ανάλυσης. Αν τα δεδομένα είναι ασυσχέτιστα δεν έχει νόημα να συνεχίσουμε. Απόλυτες τιμές του συντελεστή συσχέτισης μεγαλύτερες του 0.3 είναι ικανοποιητικές.
- Η ορίζουσα του δειγματικού πίνακα συσχέτισης είναι μεγαλύτερη από 0.00001
- Έλεγχος της υπόθεσης της σφαιρικότητας (Bartlett's test of sphericity)

# Bartlett's test of sphericity

Η προς έλεγχο υπόθεση είναι:

$$H_0 : R = I_p$$

ή

$$H_0 : \Sigma = \sigma^2 I_p$$

Το στατιστικό είναι:

$$L = - \left[ -n - \frac{1}{6(2p+5)} \right] \ln |R| \stackrel{H_0}{\sim} X_{p(p-1)/2}^2$$

ή

$$L = - \left[ -n - \frac{(2p^2 + p + 2)}{6p} \right] \left[ \ln |S| - \ln \prod_{i=1}^p s_i^2 \right] \stackrel{H_0}{\sim} X_{p(p-1)/2}^2$$

# Καταλληλότητα μεταβλητών

- Εξέταση της πολυσυγγραμμικότητας.
  1. Χρησιμοποιώντας το στατιστικό των Kaiser Meyer Olkin (ΚΜΟ) εξετάζεται αν τα δεδομένα μας είναι κατάλληλα για παραγοντική ανάλυση. Το στατιστικό αυτό παίρνει τιμές στο διάστημα  $[0,1]$ . Αν  $KMO > 0.6$  συνεχίζουμε την παραγοντική ανάλυση.

# Καταλληλότητα μεταβλητών

2. Κοιτούμε τον δείκτη Measure of Sampling Adequacy για να αποφανθούμε αν μία μεταβλητή είναι κατάλληλη για να χρησιμοποιηθεί στην ανάλυση. Τιμές μεγαλύτερες του 0.5 μας υποδεικνύουν την καταλληλότητα.

# Επιλογή αριθμού παραγόντων

1. Kaiser criterion: ο αριθμός των παραγόντων είναι ίσος με τον αριθμό των ιδιοτιμών του πίνακα συσχέτισης που είναι μεγαλύτερες από τη μονάδα.
2. Scree plot test: ο αριθμός των παραγόντων προσδιορίζεται από το γράφημα των ιδιοτιμών του πίνακα συσχέτισης σε φθίνουσα σειρά. Είναι ίσος με το πλήθος των ιδιοτιμών πριν την τελευταία σημαντική πτώση του μεγέθους της ιδιοτιμής. ( καλό για  $n > 200$ )

# Μέθοδοι εκτίμησης των παραγόντων

- Principal component analysis
- Maximum likelihood factoring
- Principal axis factoring
- Image factoring
- Alpha factoring
- Unweighted least squares factoring
- Generalizes least squares factoring

# Principal component analysis

Η πιο ευρέως χρησιμοποιούμενη μέθοδος. Αναζητεί το γραμμικό συνδυασμό των μεταβλητών έτσι ώστε να επεξηγείται από τους παράγοντες η μέγιστη μεταβλητότητα των μεταβλητών.



# Οι υπόλοιπες μέθοδοι συνοπτικά

- Principal axis factoring: παραλλαγή της PCA. Χρησιμοποιείται όταν ο πίνακας συσχέτισης είναι ιδιάζων.
- Image factoring: εφαρμόζεται στην θεωρία εικόνας.
- Unweighted least squares: ελαχιστοποιεί το άθροισμα τετραγώνων μεταξύ παρατηρούμενων και εκτιμώμενων πινάκων συσχέτισης.
- Generalized least squares: απλή παραλλαγή της Unweighted least squares

# Οι υπόλοιπες μέθοδοι συνοπτικά

- Alpha factoring: προσπαθεί να μεγιστοποιήσει την αξιοπιστία των παραγόντων
- Maximum Likelihood method: προϋποθέτει την ύπαρξη πολυδιάστατης κανονικής κατανομής. Ίσως οδηγεί σε πολλούς παράγοντες άλλα όχι περισσότερους από  $\lfloor p/2 \rfloor$ .

# Περιστροφή

- Χρησιμοποιείται για να γίνουν τα αποτελέσματά μας πιο ερμηνεύσιμα.
- Ελπίζουμε ότι οι επιβαρύνσεις κάποιων παραγόντων θα είναι μεγάλες σε απόλυτη κλίμακα μόνο για κάποιες μεταβλητές και έτσι βλέποντας ποιες μεταβλητές εξαρτώνται με ποιους παράγοντες μπορούμε να τους ερμηνεύσουμε.

# Μέθοδοι περιστροφής

1. Varimax: ελαχιστοποιεί αριθμό μεταβλητών που έχουν μεγάλες επιβαρύνσεις για κάθε παράγοντα.
2. Quartimax: ελαχιστοποιεί αριθμό παραγόντων που εξηγούν μία μεταβλητή
3. Equimax: συνδυασμός των άνω
4. Direct oblimin rotation
5. Promax

Οι 4,5 δίνουν συσχετισμένους παράγοντες.

# Υπολογισμός των σκορ των παραγόντων

- Χρησιμοποιούνται ως νέες μεταβλητές για τη διεξαγωγή περαιτέρω έρευνας.

$$F_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p$$

$$F_2 = a_{21}X_1 + a_{22}X_2 + \dots + a_{2k}X_p$$

$$F_k = a_{p1}X_{11} + a_{p2}X_2 + \dots + a_{pk}X_p$$

- Κύριες συνιστώσες έχει ως αποτέλεσμα παράγοντες ακριβείς, αλλιώς προσεγγίζονται με Regression ή Bartlett ή Anderson μέθοδο.

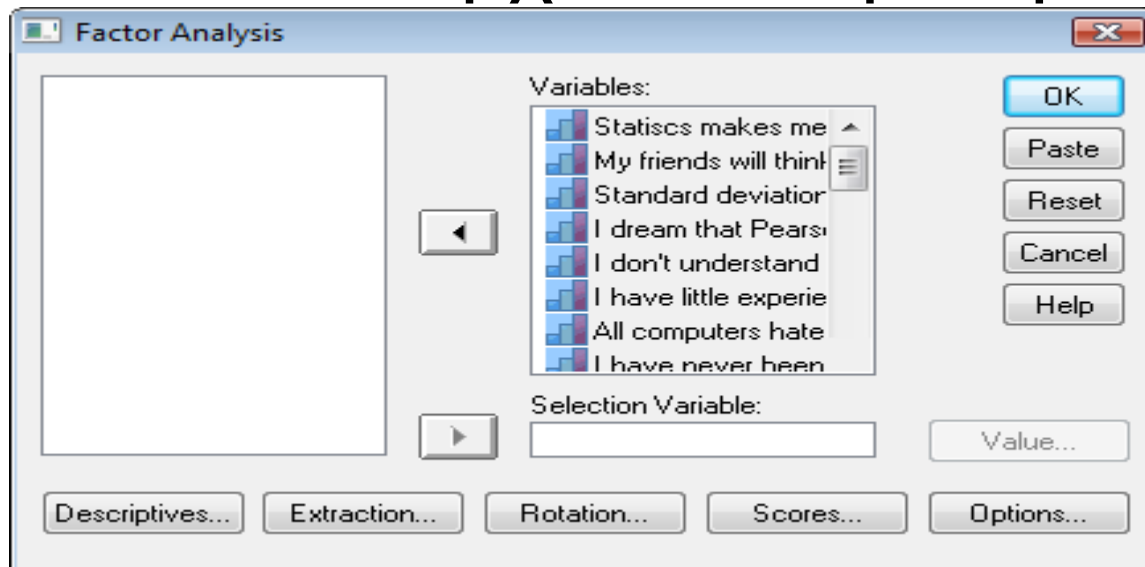
# Παράδειγμα

## Andy Field (10-12-2005)

- Στο αρχείο SAQ.sav καταγράφονται οι απαντήσεις σε 23 πιθανές ερωτήσεις για το άγχος των φοιτητών για το SPSS. Το ενδιαφέρον αρχικά επικεντρώνεται στην εύρεση πιθανών παραγόντων αιτιολόγησης του άγχους.

# Υλοποίηση στο SPSS

- Analyze>Data Reduction>Factor
- Στο νέο παράθυρο διαλόγου στο πλαίσιο Variables υπεισέρχονται οι μεταβλητές



Selection variable: αν θέλουμε να επιλέξουμε για την παραγοντική ανάλυση μόνο πειραματικές μονάδες με αυτή την τιμή για την επιλεγμένη μεταβλητή



# Descriptives

## Πλαίσιο Statistics

1. Univariate descriptives: μέση τιμή, τυπική απόκλιση και πλήθος πειραματικών μονάδων για κάθε μεταβλητή.
2. Initial solution μας δίνονται οι αρχικές ιδιοτιμές και το ποσοστό της μεταβλητότητας που εξηγείται.

# Descriptives

## Πλαίσιο Correlation Matrix

1. **Coefficients:** ο πίνακας συσχέτισης.
2. **Significance level:** δίνονται οι p-τιμές για τον έλεγχο τις συσχέτισης κάθε ζεύγους μεταβλητών.
3. **Determinant:** η ορίζουσα του πίνακα συσχέτισης. Πρέπει να είναι μεγαλύτερη από 0.00001
4. **KMO and Bartlett test of sphericity.**

# Descriptives

## Πλαίσιο Correlation Matrix

5. Inverse

6. Reproduced: ο εκτιμώμενος πίνακας συσχέτισης. Δίνονται και τα υπόλοιπα. Θα πρέπει να έχουν απόλυτες τιμές μικρότερες του 0.05

7. **Anti-image:** δίνονται οι συντελεστές μερικής συσχέτισης με αλλαγμένα πρόσημα στα μη διαγώνια στοιχεία, ενώ στα διαγώνια δίνεται ο δείκτης MSA (θέλουμε τιμές μεγαλύτερες του 0,5)

# Extraction

- Επιλέγουμε τη μέθοδο Principal Components
- Correlation matrix: για να χρησιμοποιηθεί ο πίνακας συσχετίσεων.
- Unrotated factor solution
- Scree plot
- Extract: είτε καθορίζουμε τον αριθμό των παραγόντων είτε Kaiser Criterion

# Rotation

- Επιλέγουμε Varimax.
- Display : Rotated Solution και Loading plots. Με αυτό τον τρόπο μας δίνεται η λύση με την περιστροφή και το 3-D γράφημα των πρώτων τριών παραγόντων. Αν οι παράγοντες είναι δύο μας δίνεται ένα 2-D.

# Scores

- Saves as variable: δημιουργείται (με μία εκ των τριών διαθέσιμων μεθόδων) μία νέα μεταβλητή για κάθε παράγοντα.
- Display factor score coefficient matrix: μας δίνονται οι συντελεστές με τους οποίους κάθε μεταβλητή πολλαπλασιάστηκε για τη δημιουργία των factor scores.

# Options

Καθορίζουμε

- A) τον τρόπο χειρισμού των ελλιπών τιμών
- B) τον τρόπο εμφάνισης των συντελεστών
- Γ) να μην εμφανίζονται συντελεστές με τιμές μικρότερης από αυτήν που καθορίζεται (συνήθως σε έρευνες των ανθρωπιστικών σπουδών 0.3-0.4)

# Αποτελέσματα -Ερμηνεία

- Δεν υπάρχει πρόβλημα συσχετισμένων μεταβλητών.
- Δεν υπάρχει πρόβλημα πολυσυγγραμμικότητας.
- Η Π.Α είναι κατάλληλη ( $KMO > 0.6$ ) και p-τιμή του Bartlett's test of sphericity  $< 0.05$ .



# Αποτελέσματα -Ερμηνεία

- Αριθμός παραγόντων με το κριτήριο του Keiser είναι ίσος με 4. Οι παράγοντες αυτοί εξηγούν το 50.317% της μεταβλητότητας
- Από το Scree plot θα λέγαμε είτε 4 είτε 2 παράγοντες.
- Component Matrix: τι μας υποδηλώνεται σε ένα τέτοιο πίνακα;

# Component matrix

	F_1	F_2	F_3	F_4
X_1	L_11	L_12	L_13	L_14
X_2	L_21	L_22	L_23	L_24
X_p	L_p1	L_p2	L_p3	L_p4

# Rotation

- Κοιτούμε το Rotated Component matrix. Για διευκόλυνση στην ερμηνεία στο πλαίσιο `Options>sorted by size>suppress absolute values less than 0.4`
- Τώρα είμαστε σε θέση να πούμε από ποιες μεταβλητές αποτελείται ο κάθε παράγοντας.

# Όνομασία παραγόντων

1. Φόβος για υπολογιστές
2. Φόβος για στατιστική
3. Φόβος για μαθηματικά
4. Φόβος σε σχέση με τους άλλους.

# Διάφοροι πίνακες του Output

- Component transformation matrix: ο πίνακας αυτός χρησιμοποιήθηκε για να προκύψει από τον Component matrix ο Rotated component matrix.
- Component score coefficient matrix: ο πίνακας με τα σκορ των παραγόντων
- Component score covariance matrix: ο πίνακας συνδιακύμανσης των σκορ.

# Βιβλιογραφία

- Coakes and Steed (1999). Spss analysis without anguish. Chapter 15.
- David Garson. Reliability analysis. [www2.chass.ncsu.edu/garson/PA765/reliab.htm](http://www2.chass.ncsu.edu/garson/PA765/reliab.htm)
- DeCoster, J. (2005). Overview of factor analysis. [www.stat-help.com/notes.html](http://www.stat-help.com/notes.html)

# Βιβλιογραφία

- Δ. Καρλής(2005). Πολυμεταβλητή στατιστική ανάλυση.
- Δημητριάδης (2003). Στατιστικές εφαρμογές με SPSS.
- Andy Field (2005). Factor Analysis using SPSS.  
<http://www.statisticshell.com/statlist.html>