

Çok Değişkenli Analiz Yöntemleri

Doç.Dr. Ayla ÖZHAN DEDEOĞLU
Ege Üniversitesi İİBF

Simpson Paradoxu

- Erkek adayların %44'ü kabul edilirken, kadınların sadece 33%'ü kabul edilmiştir.
- Bu ayrımcılık anlamına gelir mi?

	Erkek	Kadın	Toplam
Kabul	35	20	55
Sütun yüzdesi	44%	33%	
Satır yüzdesi	64%	36%	100%
Red	45	40	85
Sütun yüzdesi	56%	67%	
Satır yüzdesi	53%	47%	100%
Toplam	80	60	140

Simpson Paradoksu

- Cinsiyet ve programa kabul arasında ilişki yoktur.
- Neden?
 - Kadınlar dil bölümlerine daha fazla başvurmaktadır ama programa kabul zordur
 - Erkekler mühendislik bölümlerine daha fazla başvurmaktadır ve bu bölümlerin kabul oranı daha yüksektir.

Mühendislik	Erkek	Kadın	Toplam
Kabul	30	10	40
Sütun yüzdesi	50%	50%	
Satır yüzdesi	75%	25%	100%
Red	30	10	40
Sütun yüzdesi	50%	50%	
Satır yüzdesi	75%	25%	100%
Toplam	60	20	80

Filoloji	Erkek	Kadın	Toplam
Kabul	5	10	15
Sütun yüzdesi	25%	25%	
Satır yüzdesi	33%	67%	100%
Red	15	30	45
Sütun yüzdesi	75%	75%	
Satır yüzdesi	33%	67%	100%
Toplam	20	40	60

Simpson Paradoksu

- İki deęişkenli (bivariate) analiz (çapraz tablo ya da korelasyon) yanlış yönlendirebilmektedir.
- Yeni bir deęişken eklendiğinde verileri daha iyi anlamak mümkün
 - İlk yorumlar tersine bile dönebilir.

■ Arařtırmalarda genelde ok sayıda iliřkili deęiřkenler sz konusu

■ Genelde arařtırmacılar frekans analizleri ve apraz tablo ile yetinmektedirler.

■ Ancak bu analizler yanlış ynlendirebilir.

■ "Numbers don't lie, you can lie with numbers"

Çok Değişkenli Analizler ve Varsayımları

■ Çok sayıda bağımsız ve bağımlı değişken arasında eş anlı analizi yapılır.

■ Varsayımlar

- Veriler çok değişkenli normal dağılıma uyar.
- Tüm gruplar için kovaryans matrisleri eşittir.
- Bağımsız değişkenler arasında anlamlı doğrusal ilişki yoktur.
- Bağımlı ve bağımsız değişkenler arasındaki ilişki doğrusaldır.
- Değişken birim değerleri birbirinden bağımsızdır

Ölçekler

■ Dört düzey

1. *İsimsel (Nominal) ölçek*: Kategorik. örn. cinsiyet
0=kadın, 1=erkek
2. *Sıralı (Ordinal) ölçek*: Sıralama (en küçük=1 en büyük=10)
3. *Aralıklı (Interval) ölçek*: Sürekli. örn. sıcaklık derecesi, yaş grupları
4. *Oransal (Ratio) ölçek*: 0 noktası olur. (yükseklik, yaş)

■ Not: 1 & 2= Metrik olmayan veri
3 & 4 = Metrik veri

Ö
l
ç
e
ğ
i
n
g
ü
c
ü

Ölçekler ve Analizler Arasındaki İlişki

- Ölçekler, verilerin istatistiksel olarak işlenmesine temel oluşturur.
- Temel değişkenin ölçeğine göre test seçilir.
 - Metrik veriler için parametrik testler
 - Metrik olmayanlar için nonparametrik testler
- NOT: Nominal ve ordinal değişkenler için ortalama ve std.sapma hesaplanmaz Parametrik testler yapılamaz.

Tek deęişkenli Analizler (Metrik ölçekler)

■ Tek örneklem

- T test

■ İki ya da daha fazla örneklem

■ Bağımsız

- İki gruplu t-testi
- One-way ANOVA

■ İlişkili

- İlişkili örneklem t testi

Tek deęişkenli Analizler (Metrik olmayan ölçekler)

■ Tek örneklem

- Frekans
- Ki-kare
- K-S
- Runs
- Binomial

■ İki ya da daha fazla örneklem

■ Baęımsız

- Ki-kare
- Mann-Whitney
- Median
- K-S

■ İlişkili

- Sign
- Wilcoxon
- McNemar
- Ki-kare

Çok Değişkenli Analizler

■ 2 genel tip

■ Bağımlılık analizi

- Bir ya da daha fazla değişkenin bağımlı değişken olduğu ve diğer değişkenlerce açıklandığı ve tahmin edildiği durumlarda
 - Örnek. Çoklu regresyon

■ Bağımsızlık analizi

- Hiç bir değişken "bağımlı" değildir.
- Değişkenler ya da vakalar arasındaki ilişkiye bakılır.
 - Örnek. Kümeleme analizi, faktör analizi

Bağımlılık Analizleri

- Çoklu Regresyon
- Ayırma Analizi (Discriminant Analysis)
- Logit/Lojistik Regresyon
- Multivariate Analysis of Variance (MANOVA) ve Kovaryans
- Conjoint Analizi
- Kanonik Korelasyon
- Yapısal Eşitlik Modeli (Structural Equations Modeling (SEM))

Bağımsızlık Analizleri

- Asal Bileşenler (Principal Components) ve Factor Analizi
- Kümeleme Analizi (Cluster Analysis)
- Çok Boyutlu Ölçekleme (Multidimensional Scaling)
- Correspondence Analizi

Çok Değişkeli Analiz Tekniği Seçme

1. Araştırılan ilişkinin tipi nedir?

Bağımlılık - bağımsızlık?

2. Bağımlılık İlişkisi: Kaç değişken tahmin edilecek?

✓ Bağımlı değişkenin ölçeği nedir?

✓ Tahmin edici değişken(ler)in ölçeği nedir?

3. Bağımsızlık İlişkisi: Değişkenler, katılımcılar ya da nesnelere arasındaki ilişkiyi mi inceliyoruz?

Teknikler

- Çoklu regresyon; tek bir sürekli değişken (bağımlı değişken), 2 ve daha çok sürekli ya da nominal değişken (bağımsız ya da tahmin edici değişkenler) tahmin edilmeye çalışılır.

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

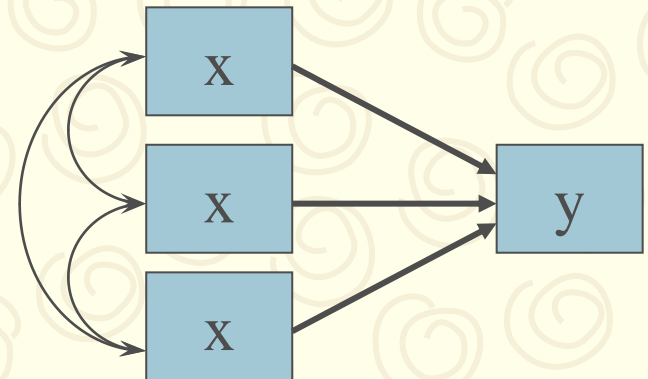
Metrik metrik , metrik olmayan

Y : Bağımlı değişken

X_1, X_2, X_3, \dots : Bağımsız değişkenler

α, β : Katsayılar

u_i : Hata terimi



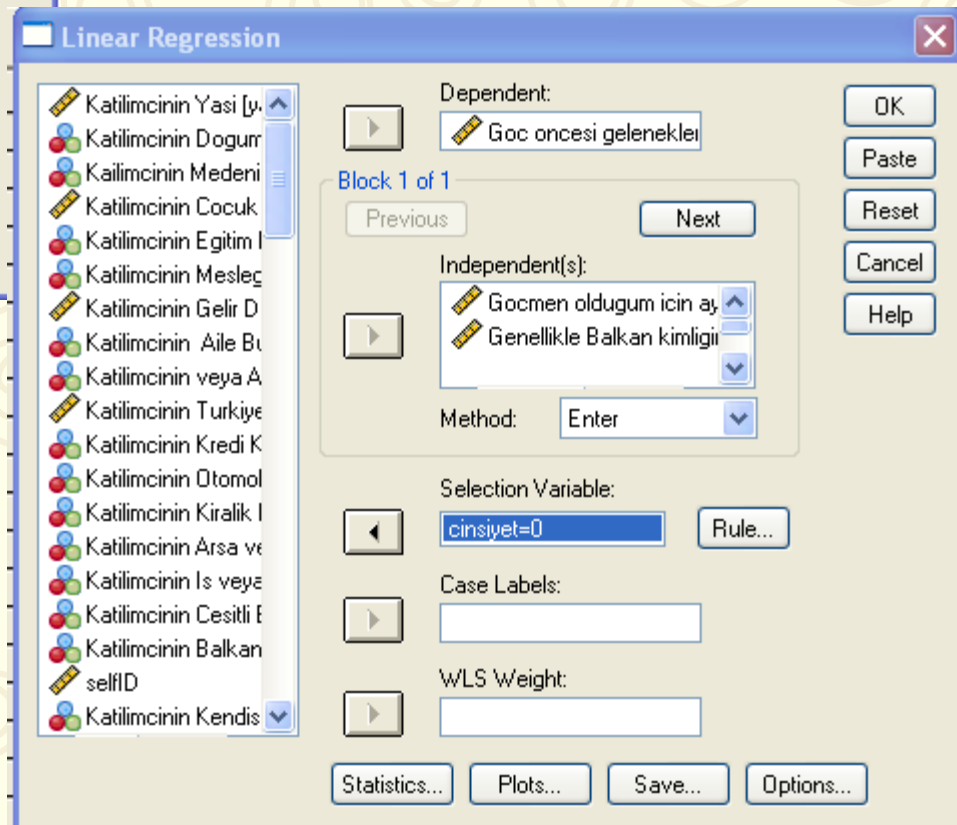
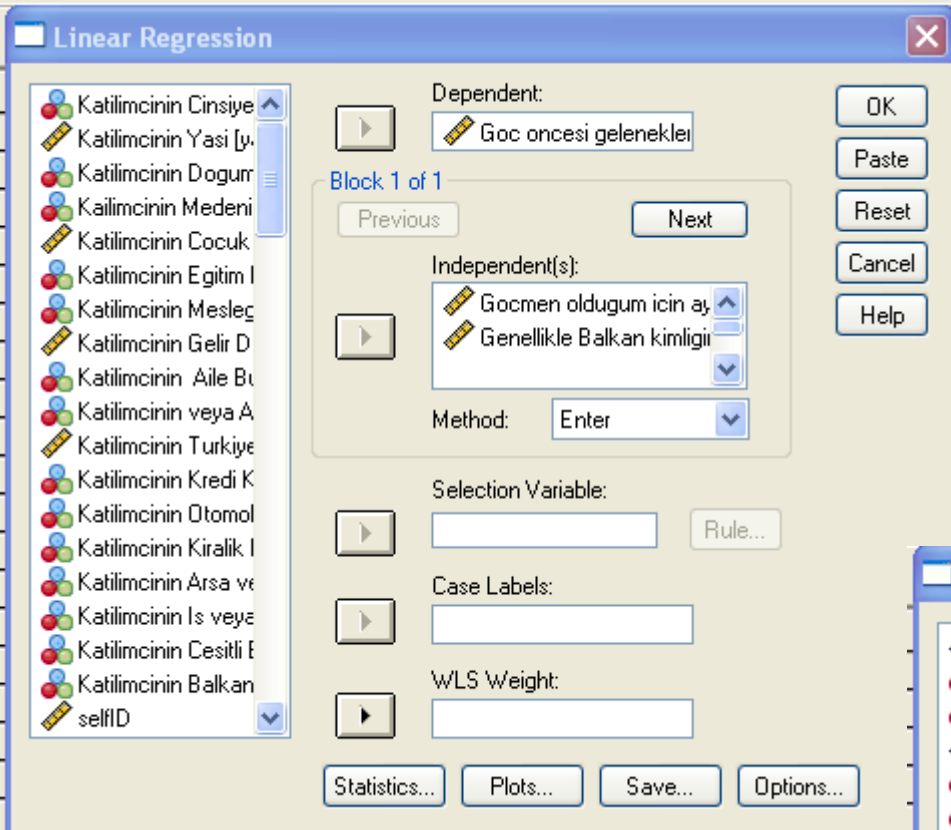
Regresyon

- Aşamalı regresyon analizinde amaç, bağımsız değişken sayısını azaltarak en önemli tahminleyicileri belirlemek

$$y_i = \alpha + \beta_1 x_{1i} + u_i$$

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$$

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,295 ^a	,087	,082	1,119

- a. Predictors: (Constant), Genellikle Balkan kimligimi hatirlatan veya on plana cikaran urunleri ve markalari kullaniyorum, Gocmen oldugum icin ayrimciliga maruz kaldigimi dusunuyorum

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	45,301	2	22,650	18,102	,000 ^a
	Residual	474,218	379	1,251		
	Total	519,518	381			

- a. Predictors: (Constant), Genellikle Balkan kimligimi hatirlatan veya on plana cikaran urunleri ve markalari kullaniyorum, Gocmen oldugum icin ayrimciliga maruz kaldigimi dusunuyorum
- b. Dependent Variable: Goc oncesi geleneklerimizi buyuk olcude korudugumu dusunuyorum

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3,092	,175		17,636	,000
	Gocmen oldugum icin ayrimciliga maruz kaldigimi dusunuyorum	-,165	,048	-,179	-3,478	,001
	Genellikle Balkan kimligimi hatirlatan v eya on plana cikaran urunleri ve markalari kullaniyorum	,305	,053	,294	5,722	,000

a. Dependent Variable: Goc oncesi geleneklerimizi buyuk olcude korudugumu dusunuyorum

$$y = 3,092 - 0,165x_1 + 0,305x_2 + u$$

Lojistik Regresyon

- Bağımlı değişken 0 ve 1 gibi ikili (binary) ya da ikiden çok kategori içeren kesikli kategorik değişken olduğunda normallik varsayımı bozulmakta ve doğrusal regresyon analizi uygulanamamaktadır.
- Tahmin edici değişkenler kalitatif ve kategorik (örn. cinsiyet) ya da kantitatif (örn. test değerleri) olmalı
- Bağımlı değişken kategorik olmalı
 - Bağımlı değişken sürekli ise, çoklu regresyon kullanılmalıdır.

Lojistik Regresyon

- Bağımlı ve bağımsız değişken arası ilişki doğrusal değilse

$$L = \ln(P / 1 - P) = \alpha + \beta_1 x_1 + \beta_2 x_2 + u$$

Metrik olmayan

metrik + metrik olmayan

İki gruplu ise (binary logistic)

İkiden fazla gruplu ise (multinomial logistic)

Tek kategorik bağımsız değişkenli lojistik regresyon örneği

$$L = \ln(P / 1 - P) = \alpha + \beta_1 x_{1i}$$

	finansaldurum	büyüklik	vs
1	1	1	
2	1	0	
3	1	1	
4	1	1	
5	1	1	
6	1	1	
7	1	0	
8	1	0	
9	1	1	
10	0	0	
11	0	0	
12	0	0	
13	0	0	
14	0	0	
15	0	1	
16	0	0	
17	0	1	
18	0	0	
19			

Finansal durum

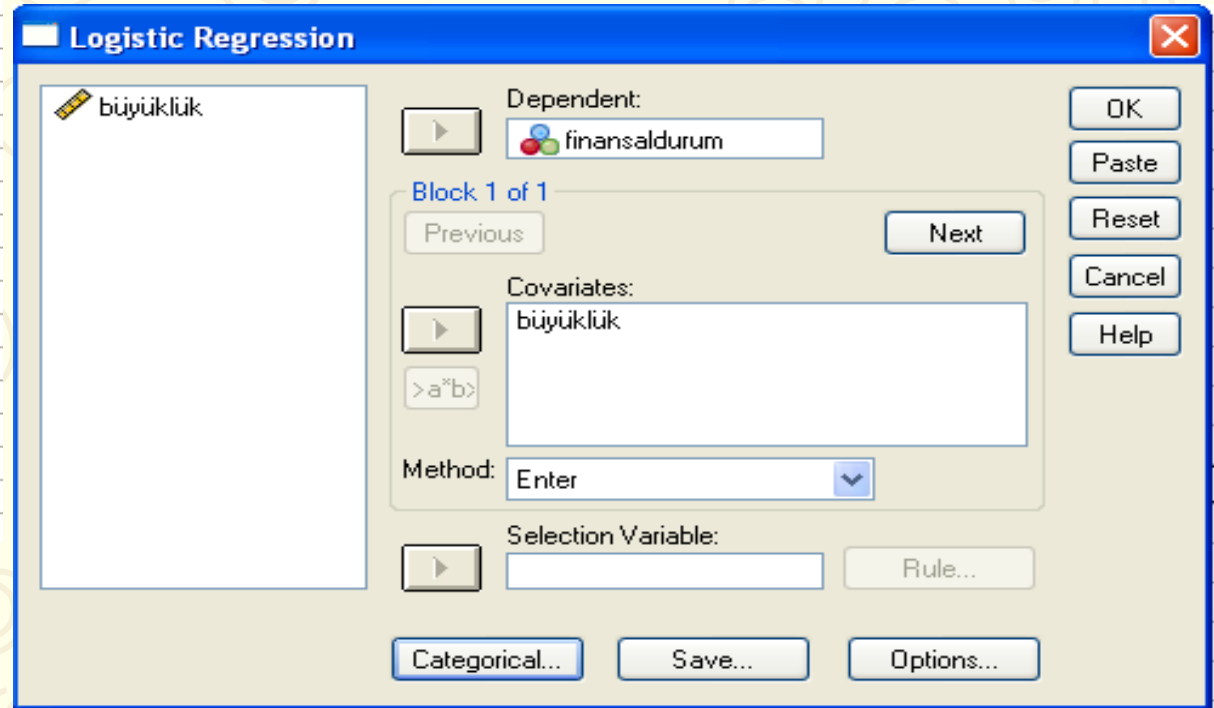
0: iflas etmiş

1: iflas etmemiş

Büyüklik

0: Küçük

1: Büyük



Albayrak vd. 2005, SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri, Asil Yayın: Ankara.

■ ML (maximum likelihood) istatistiği logistic regresyonda $-2\log L$

■ Bağımlı değişkendeki açıklanmayan varyansın anlamlılığını gösterir → Anlamli olmaması istenir.

■ Modelin verileri tam temsil etmesi durumunda değeri 0 olur.

1. İlk $-2\log L$ sadece sabit terimli denklem için hesaplanır.

2. İkincisi yeni bir değişken eklenerek hesaplanır.

3. Aradaki fark yeni değişkenin modele katkısını gösterir.

- İlk iterasyonda 24,953, ikincide büyüklük değişkeni eklenerek 21,215
- İki $-2\log L$ arasındaki fark anlamlı değil. $p > 0,05$ - Model verileri temsil etmiyor

Iteration History^{a,b,c}

Iteration	-2 Log likelihood	Coefficients	
		Constant	
Step 0	1	24,953	,000

- Constant is included in the model.
- Initial -2 Log Likelihood: 24,953
- Estimation terminated at iteration number 1 because parameter estimates changed by less than ,001.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	3,739	1	,053
	Block	3,739	1	,053
	Model	3,739	1	,053

Iteration History^{a,b,c,d}

Iteration	-2 Log likelihood	Coefficients	
		Constant	büyükük
Step 1	1	21,234	1,800
1	2	21,215	1,943
	3	21,215	1,946
	4	21,215	1,946

- Method: Enter
- Constant is included in the model.
- Initial -2 Log Likelihood: 24,953
- Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step	büyüklik	1,946	1,069	3,313	1	,069	7,000
1	Constant	-,847	,690	1,508	1	,220	,429

a. Variable(s) entered on step 1: büyüklik.

$$L = \ln(P / 1 - P) = -0,847 + 1,946 x_{1i}$$

$$P = 1 / (1 + e^{(-0,847 + 1,946x_1)}) = 1 / (1 + e^{-0,847} e^{1,946x_1})$$

$$x_1 = 1 \Rightarrow P = 0,75$$

$$x_1 = 0 \Rightarrow P = 0,30$$

- Firma büyük ise iflas etmeme olasılığı %75
- Küçük ise %30

■ Kanonik Korelasyon

- Bağımlı değişkenler seti ile tahmin edici değişkenler seti arasında ilişki

$$\begin{array}{ccc} Y_1 + Y_2 + \dots + Y_n & = & X_1 + X_2 + \dots + X_n \\ \text{metrik, nonmetrik} & & \text{metrik, nonmetrik} \end{array}$$

- Çoklu regresyon analizinin uzantısıdır.

Ayırma Analizi (Discriminant Analysis)

- Tek bir bağımlı ve kategorize değişken varsa kullanılır.
- Örneklem metrik olmayan değişkenler temelinde alt gruplara ayrılabiliriyorsa kullanılır.
- Temel amacı, (1) grup farklılıklarını açıklamak ve (2) birçok metrik bağımsız değişkene dayalı olarak bir varlığın (kişi ya da nesne) belli bir sınıf ya da gruba ait olup olmadığını tahmin etmek.

$$Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n$$

Metrik
olmayan

metrik

Kullanım amacı

- Diskriminant fonksiyonları aracılığı ile gruplar arası ayırma en fazla etki eden ayırıcı değişkenleri belirlemede
- Hangi gruptan geldiği bilinmeyen bir birimin hangi gruba dahil edileceğini belirlemede kullanılır.

Grupların kovaryans matrisleri eşit olmalı, değilse logistic regresyon kullanılabilir.

- H0: Grupların kovaryans matrisleri eşittir.
- H1: Grupların kovaryans matrisleri eşit değildir.
- Box's M=8,797 F=2,669 p>0,05

Gözlem sayısı fazla ise homojenlikten küçük sapmalar nedeniyle $p < 0,05$ çıkabilir.

Korelasyon matrisinde bağımsız değişkenler arası korelasyonlar 0,70'den büyükse değişkenlerden biri analizden çıkarılabilir ya da değişkenler birleştirilebilir.

Discriminant Analysis

Grouping Variable: cep(1 2)

Independents: acil, sohbet, ozgur1

Enter independents together
 Use stepwise method

Buttons: OK, Paste, Reset, Cancel, Help, Select >>, Statistics..., Method..., Classify..., Save...

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
ACIL	,944	23,851	1	405	,000
SOHBET	,987	5,510	1	405	,019
OZGUR1	,954	19,519	1	405	,000
SOSYALAZ	,977	9,477	1	405	,002
MODERN	1,000	,075	1	405	,785
POPULER	,997	1,165	1	405	,281
IHTIYAC	,996	1,607	1	405	,206
MESGUL	,988	4,994	1	405	,026
KUR	,977	9,453	1	405	,002
PAYLAS	,965	14,540	1	405	,000
GURUR	,989	4,455	1	405	,035
PRESTIJ	,993	2,929	1	405	,088
PERSON1	,996	1,575	1	405	,210
PERSON2	,999	,507	1	405	,477

Discriminant fonksiyonun ne kadar önemli olduğunu görmek için kanonik korelasyon, eigenvalue (özdeğer) ve Wilk's Lambda

Kanonik korelasyon = 0,437

Kanonik korelasyonun karesi, modelin bağımlı değişkendeki değişimi açıklama yüzdesini açıklar %19

Eigenvalue ne kadar büyükse bağımlı deęişkendeki varyansın daha büyük kısmı model tarafından açıklandığını gösterir. 0,40'tan büyük olmalı

Wilk's Lambda, eigenvalue istatistiğinin anlamlılığını test eder.

Denklem "canonical discriminant function coefficients" tablosundan yazılır

Ayırma analizinin başarısı doğru sınıflandırma yüzdesidir.

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	,235 ^a	100,0	100,0	,437

a. First 1 canonical discriminant functions were used in the analysis.

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	,809	83,289	22	,000

Standardized Canonical Discrimin

	Function
	1
ACIL	,455
SOHBET	-,130
OZGUR1	-,437
SOSYALAZ	,165
MODERN	-,214
POPULER	,134
IHTIYAC	-,097
MESGUL	,173
KUR	,081
PAYLAS	,141
GURUR	,188
PRESTIJ	,040

Değişkenlerin
sırası

Denklem
Katsayıları

Structure Matrix

	Function
	1
ACIL	,500
OZGUR1	-,452
PERSON3	,394
PAYLAS	,391
GOSTERIS	,357
ZIYNET	,330
SOSYALAZ	,315
KUR	,315
SOHBET	-,240
MESGUL	,229
GURUR	,216
PRESTIJ	,175

Classification Results^a

		CEP	Predicted Group Membership		Total
			1	2	
Original	Count	1	250	82	332
		2	25	50	75
	%	1	75,3	24,7	100,0
		2	33,3	66,7	100,0

a. 73,7% of original grouped cases correctly classified.

Multivariate Analysis of Variance (MANOVA)

- Bir ya da daha fazla kategorik bağımsız değişken varsa ve 2 ve daha fazla sürekli bağımlı değişken varsa kullanılır.

$$Y_1 + Y_2 + \dots + Y_n = X_1 + X_2 + \dots + X_n$$

metrik

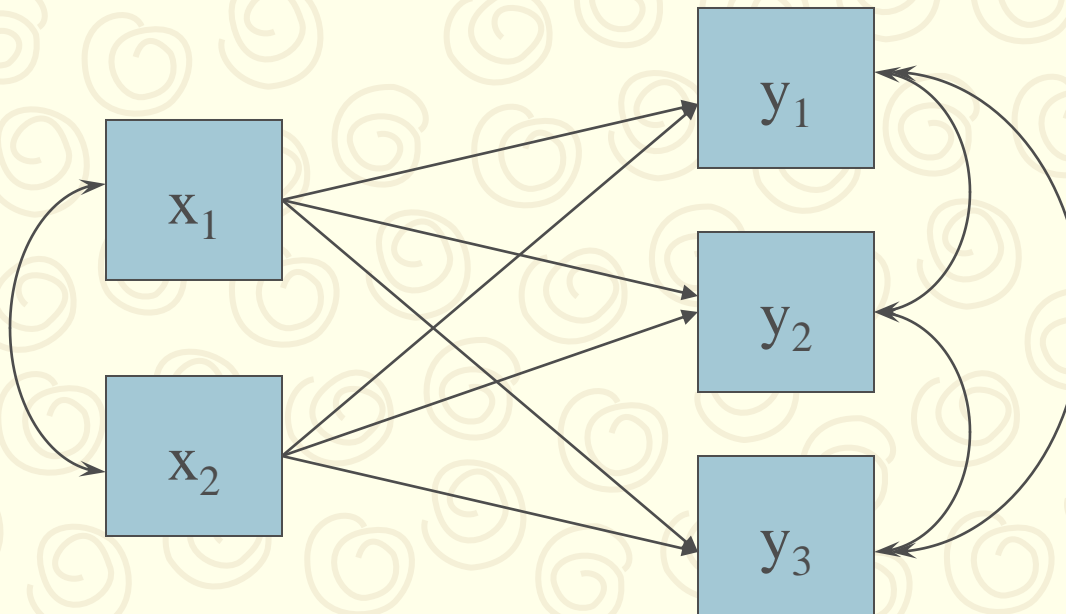
metrik olmayan, sabit faktör

metrik-covariate

■ Arařtırmacının grup cevabının varyansını dikkate alarak hipotezi test etmek için deneysel durumu tasarlamasında iki ya da daha fazla bağımsız deęişkenin, birden fazla bağımlı deęişken üzerindeki etkisi incelemede kullanılır.

MANOVA

- Tek yönlü MANOVA: Birden fazla bağımlı değişken, 1 bağımsız değişken
- Çift yönlü MANOVA: Birden fazla bağımlı değişken, 2 bağımsız değişken



Tek yönlü MANOVA

- Birden fazla bağımlı değişken, 1 bağımsız değişken

Multivariate

Dependent Variables:

- ☑ SMEAN(S664) [S6
- ☑ SMEAN(S665) [S6
- ☑ SMEAN(S666) [S6
- ☑ SMEAN(S667) [S6
- ☑ SMEAN(S668) [S6
- ☑ Ortak girisim [ortak
- ☑ Grup hakkında üçü
- ☑ Topluluk kimligi ile
- ☑ Uye isim ve profile
- ☑ Topluluk uyeliginin
- ☑ Tuketime iliskin ve
- ☑ Yuksek imajli mark
- ☑ sosyal ve topluluk
- ☑ toplulugun faaliyetl
- ☑ Muhafazakar düsu
- ☑ S5 = 3 (FILTER) [fi

Fixed Factor(s):

- ☑ Uyeliginizin niteligi sizc

Covariate(s):

WLS Weight:

Model...
Contrasts...
Plots...
Post Hoc...
Save...
Options...

OK Paste Reset Cancel Help

■ Kovaryans eşitliği şartı vardır.

■ $p > 0,05$ ise varyans eşitliği şartı sağlanmıştır.

Box's Test of Equality of Covariance Matrices

Box's M	7,657
F	1,898
df1	6
df2	84861,731
Sig.	,058

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across
a. Design: Intercept+S5

Levene's Test of Equality of Error Variances

	F	df1	df2	Sig.
Karsilikli baglilik	5,847	2	354	,053
Pay lasilan birikim	2,940	2	354	,054

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+S5

■ **Wilk's lambda** küçüldükçe bağımsız değişkenin bağımlı değişken üzerindeki etkisi artar.

■ **Diğer istatistikler** büyüdükçe bağımsız değişkenin bağımlı değişken üzerindeki etkisi artar.

Multi variate Tests^c

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	,120	24,117 ^a	2,000	353,000	,000	,120
	Wilks' Lambda	,880	24,117 ^a	2,000	353,000	,000	,120
	Hotelling's Trace	,137	24,117 ^a	2,000	353,000	,000	,120
	Roy's Largest Root	,137	24,117 ^a	2,000	353,000	,000	,120
S5	Pillai's Trace	,309	32,374	4,000	708,000	,000	,155
	Wilks' Lambda	,691	35,858 ^a	4,000	706,000	,000	,169
	Hotelling's Trace	,448	39,382	4,000	704,000	,000	,183
	Roy's Largest Root	,447	79,184 ^b	2,000	354,000	,000	,309

a. Exact statistic

b. The statistic is an upper bound on F that yields a lower bound on the significance level.

c. Design: Intercept+S5

■ **Partial Eta squared** etki düzeyini gösterir. 1'e yaklaştıkça etki düzeyi artar.

■ Üyeliğin niteliği değişkeni karşılıklı bağlılık ve paylaşılan değer değişkenini açıklamaktadır.

■ Partial eta squared incelendiğinde karşılıklı bağlılık değişkeninin etkisinin daha fazla olduğu görülür

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Karsilikli baglilik	105,848 ^a	2	52,924	74,895	,000	,297
	Pay lasilan birikim	60,298 ^b	2	30,149	36,093	,000	,169
Intercept	Karsilikli baglilik	32,502	1	32,502	45,995	,000	,115
	Pay lasilan birikim	17,948	1	17,948	21,486	,000	,057
S5	Karsilikli baglilik	105,848	2	52,924	74,895	,000	,297
	Pay lasilan birikim	60,298	2	30,149	36,093	,000	,169
Error	Karsilikli baglilik	250,152	354	,707			
	Pay lasilan birikim	295,702	354	,835			
Total	Karsilikli baglilik	356,000	357				
	Pay lasilan birikim	356,000	357				
Corrected Total	Karsilikli baglilik	356,000	356				
	Pay lasilan birikim	356,000	356				

a. R Squared = ,297 (Adjusted R Squared = ,293)

b. R Squared = ,169 (Adjusted R Squared = ,165)

Faktör analizi ve Temel bileşenler analizi

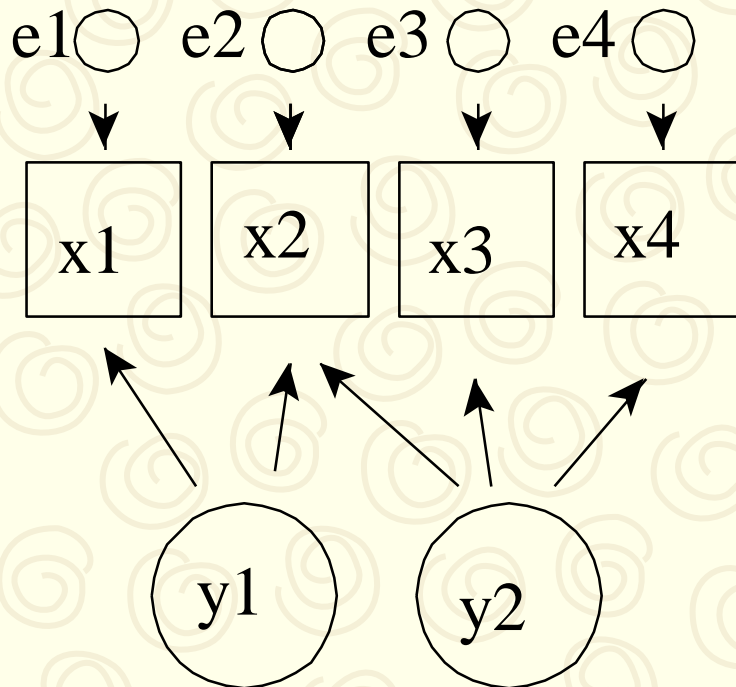
- Çok sayıda değişken arasındaki karşılıklı ilişkilerin analizinde kullanılır. Altta yatan ortak boyutlar (faktörler) temelinde, çok sayıda değişkenin daha az sayıda değişkenler setine (faktörler) en az veri kaybıyla indirgenmesine çalışılır.
- Veri tipi: Metrik değişkenler

Faktör analizi - Temel bileşen analizi

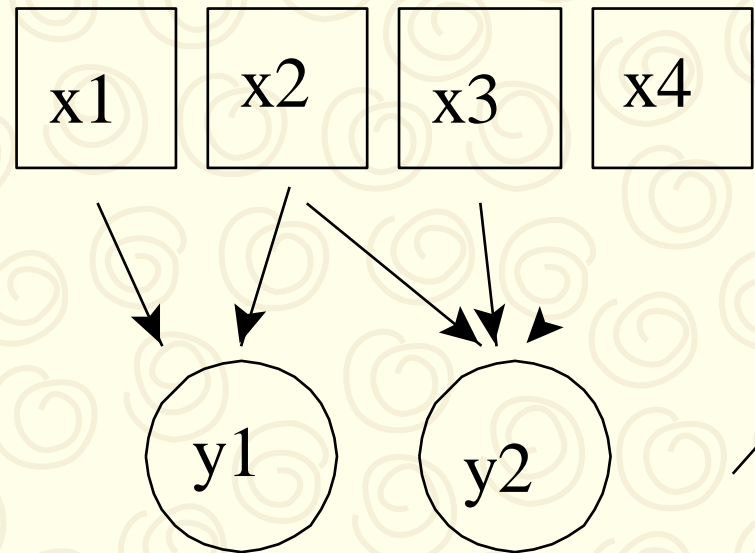
- FA verilerdeki belirleyici boyutları ortaya çıkarır ve dolayısıyla ortak varyansla ilgilidir.
- TBA ise özgün verileri bir dizi doğrusal değişken olarak kabul edip her değişkenin temel bileşene katkısını ortaya çıkarmaya çalışır

Faktör analizi - Temel bileşen analizi

FA



TBA



Faktör analizi - Temel bileşen analizi

FA

$$x1_i = a_{11}y1_i + e1_i$$

$$x2_i = a_{21}y1_i + a_{22}y2_i + e2_i$$

$$x3_i = + a_{32}y2_i + e3_i$$

$$x4_i = + a_{42}y2_i + e4_i$$

TBA

$$y1_i = a_{11}x1_i + a_{21}x2_i$$

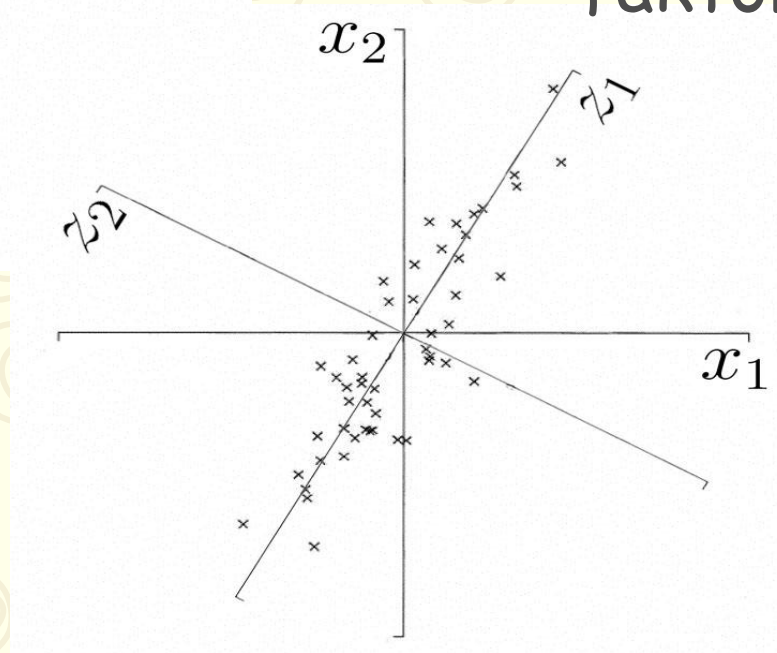
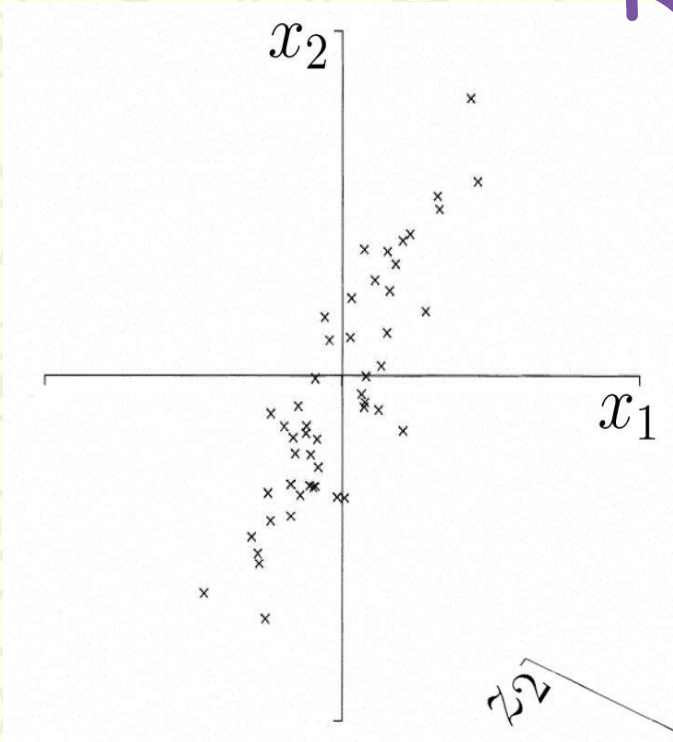
$$y2_i = a_{22}x2_i + a_{32}x3_i + a_{42}x4_i$$

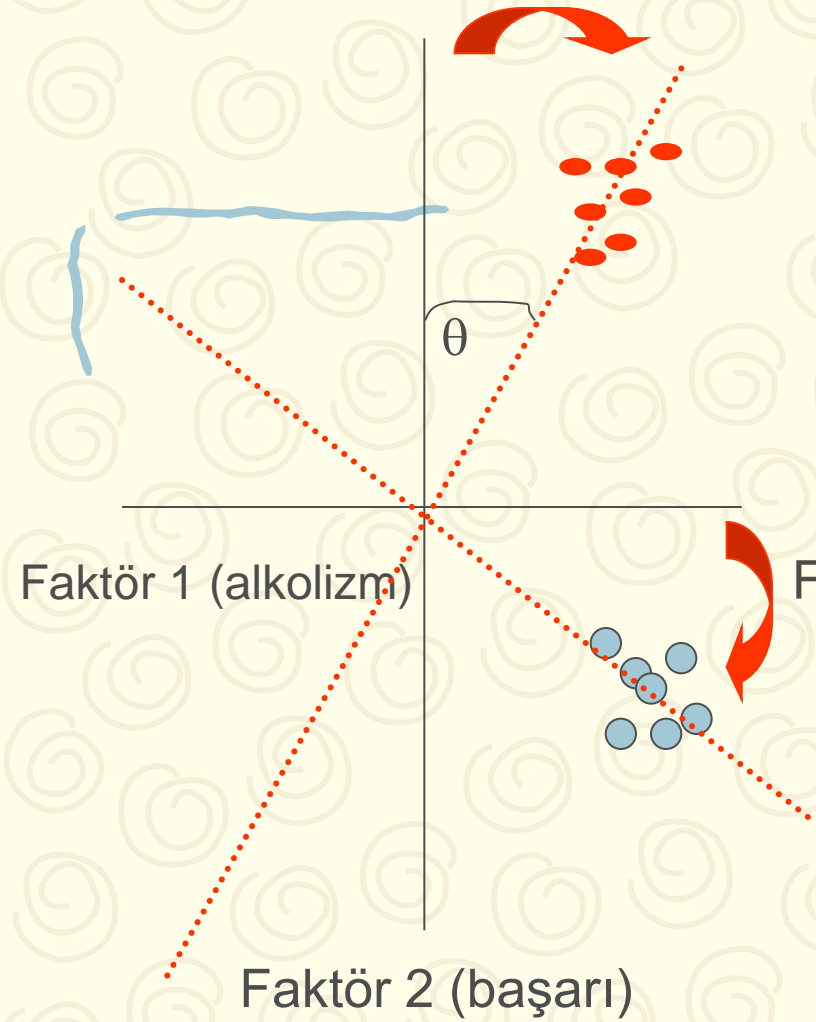
■ FA --- doğrulayıcı faktör analizi ve yapısal eşitlik modelleme (SEM). Model temellidir. Değişkenler arası ilişkileri anlama amacı.

■ TBA - Veri azaltma

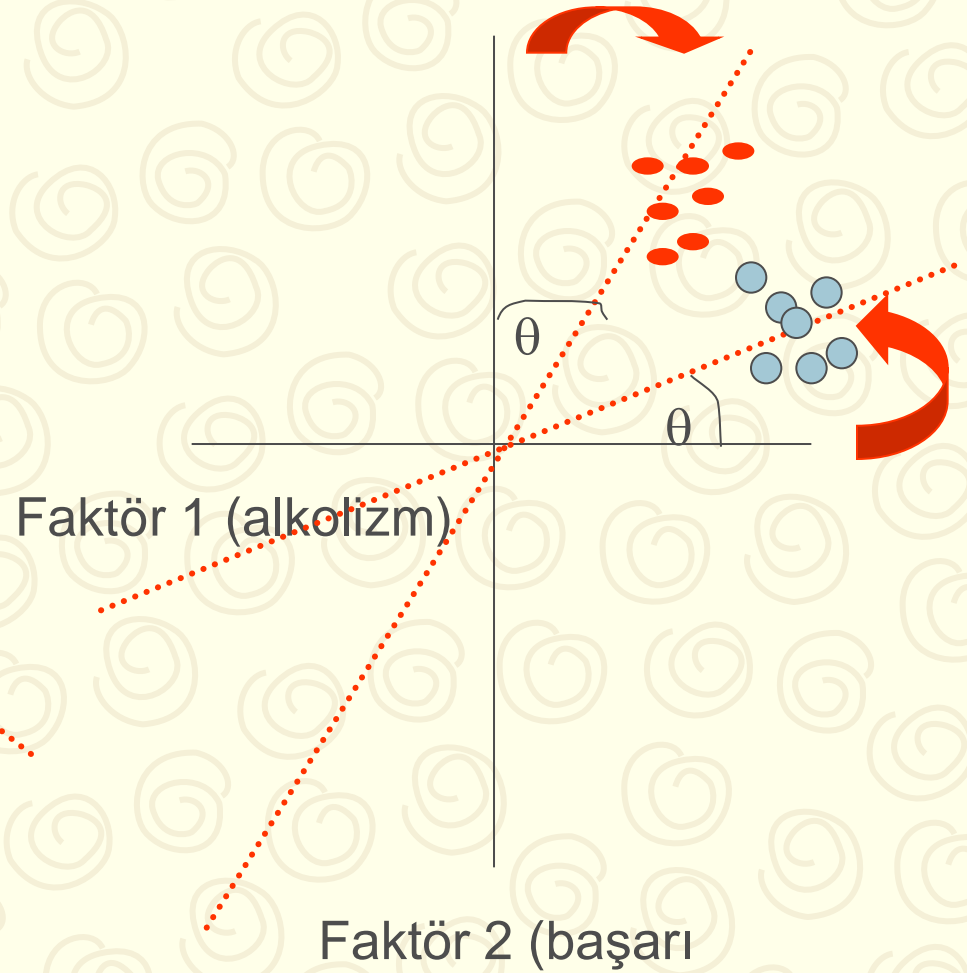
Rotasyon

- Non-orthogonal; faktörler ilişkili
- Orthogonal; faktörler ilişkisiz





İlgisiz (orthogonal) rotasyon



Eğik (oblique) rotasyon

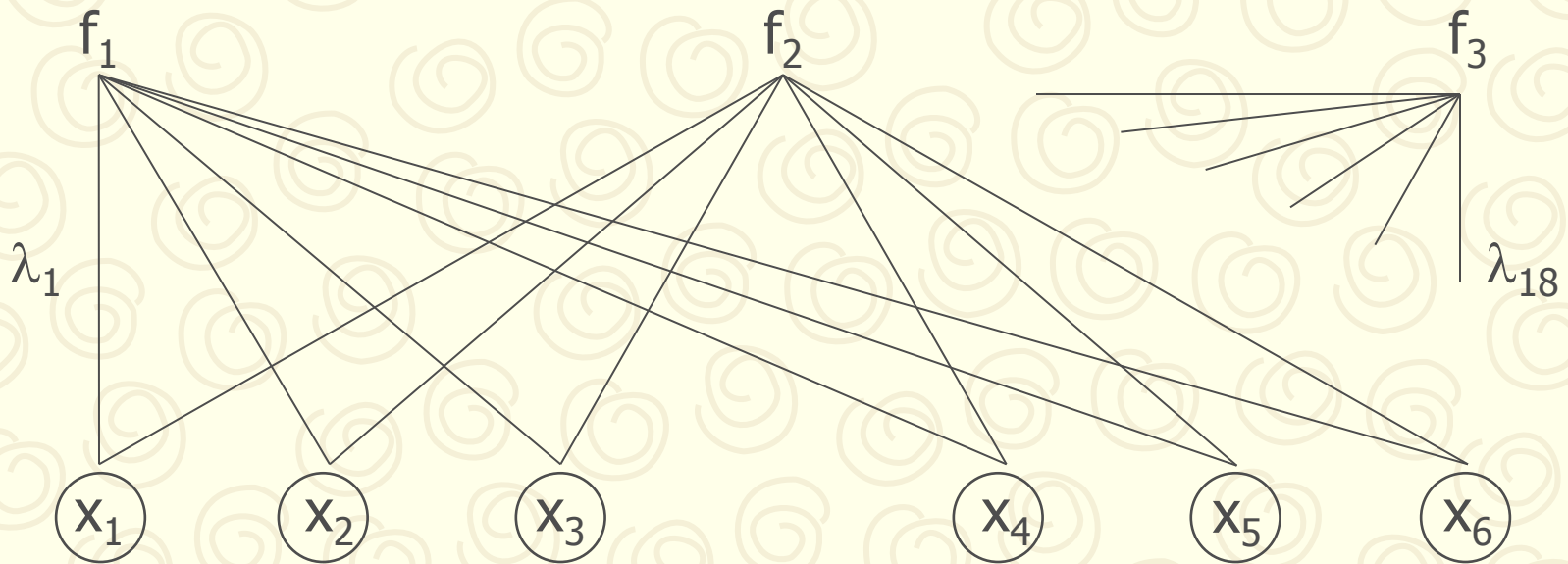
Rotasyon Tipleri

- Orthogonal; faktörler ilişkisiz
 - Varimax (faktör yorumunu kolaylaştırır)
 - Faktörlere yüklenen değişken sayısını minimize eder.
 - Quartimax (değişken yorumunu kolaylaştırır)
 - Her değişken için gerek duyulan faktör sayısını minimize eder.
 - Equamax (ikisini birleştirir)

Açıklayıcı (exploratory) ve doğrulayıcı (confirmatory) yaklaşım farkı

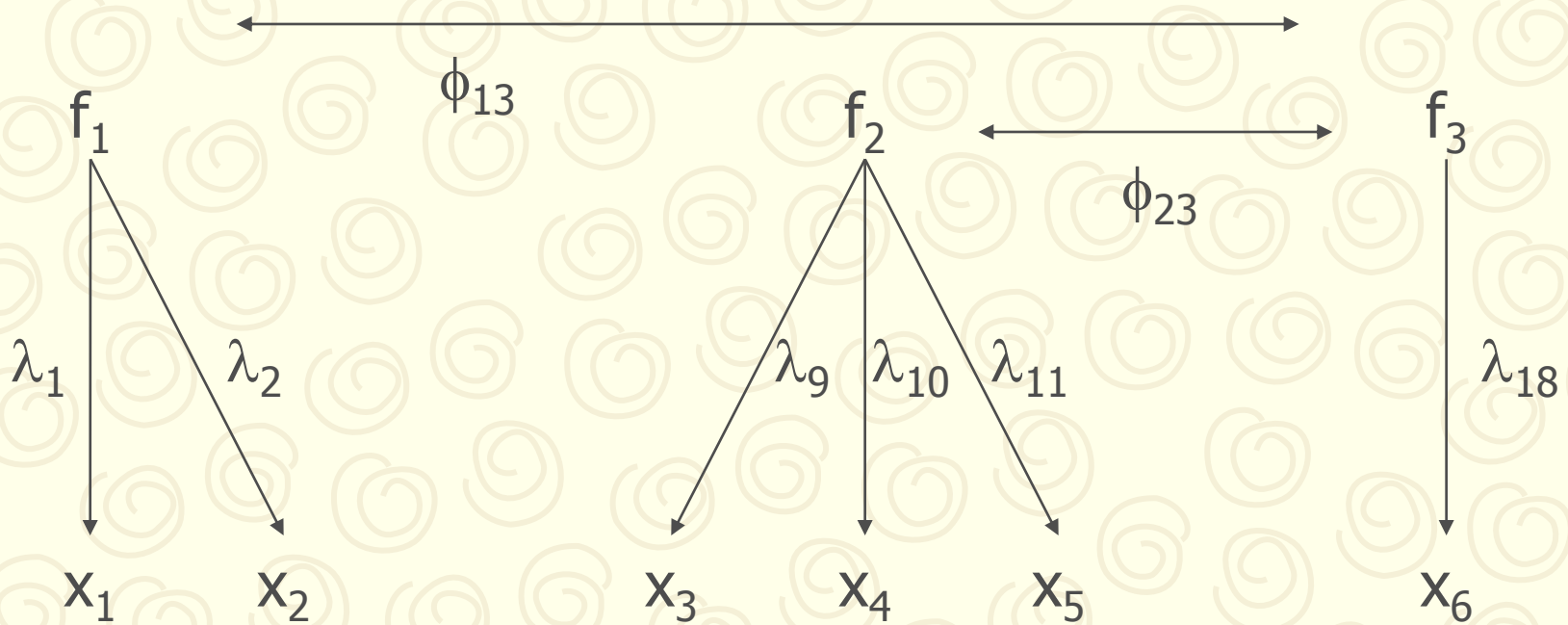
- Exploratory yaklaşımda araştırmacı yüklemeleri (loadings) kontrol edemez.
 - Amaç örneklem bulgularını genellemek
- Confirmatory yaklaşımda faktörler tahmin prosedüründe önce tanımlanır ve kavramlaştırılır.
 - Verileri incelemek veya spesifik bir hipotezi etmek

Açıklayıcı (exploratory) faktör analizi

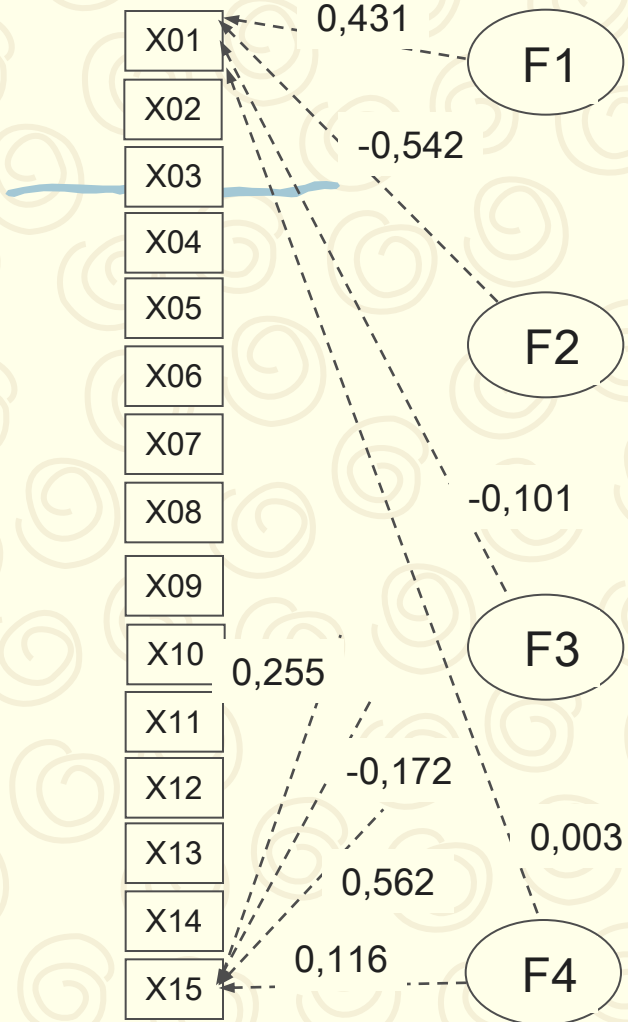


Doğrulayıcı (confirmatory) faktör analizi

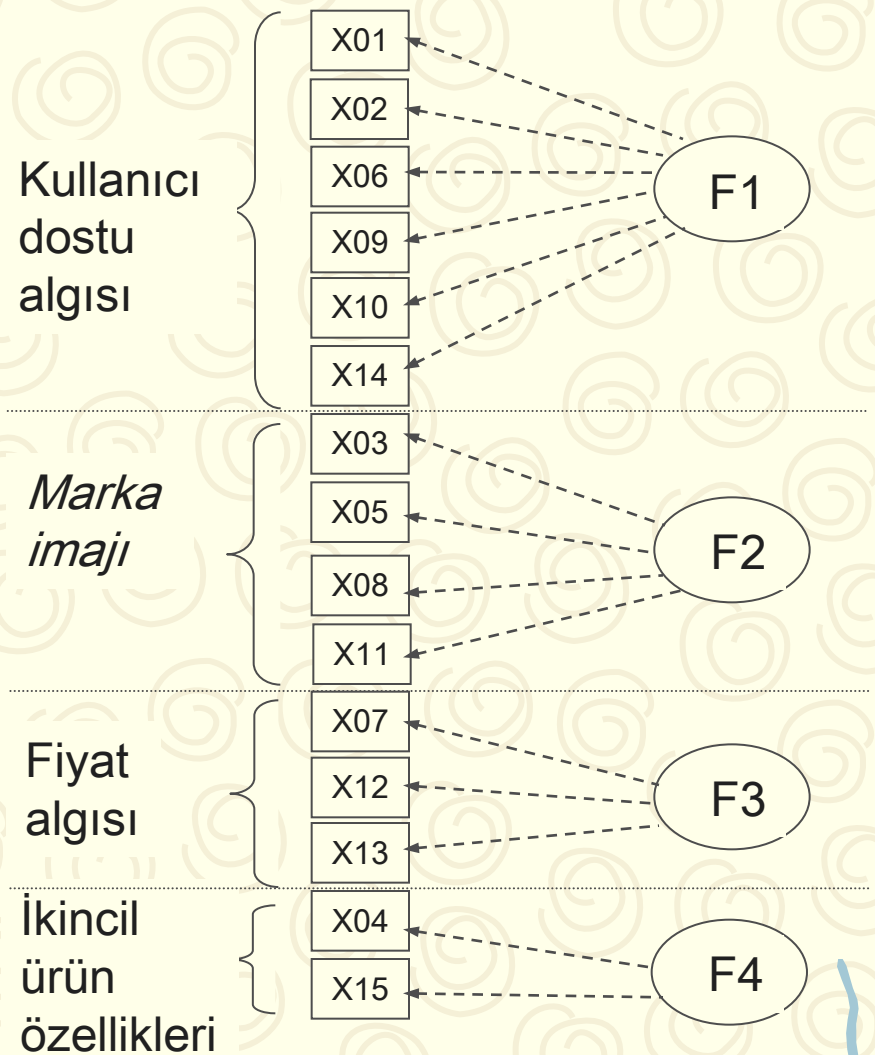
$$H_0: \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7, \lambda_8, \lambda_{12}, \lambda_{13}, \lambda_{14}, \lambda_{15}, \lambda_{16}, \lambda_{17} = 0$$



Açıklayıcı faktör analizi



Doğrulayıcı faktör analizi



15 x 4 = 60 "free" yükleme

6 + 4 + 3 + 2 = 15 önceden belirlenmiş yükleme

Çok Boyutlu Ölçekleme (MDS)

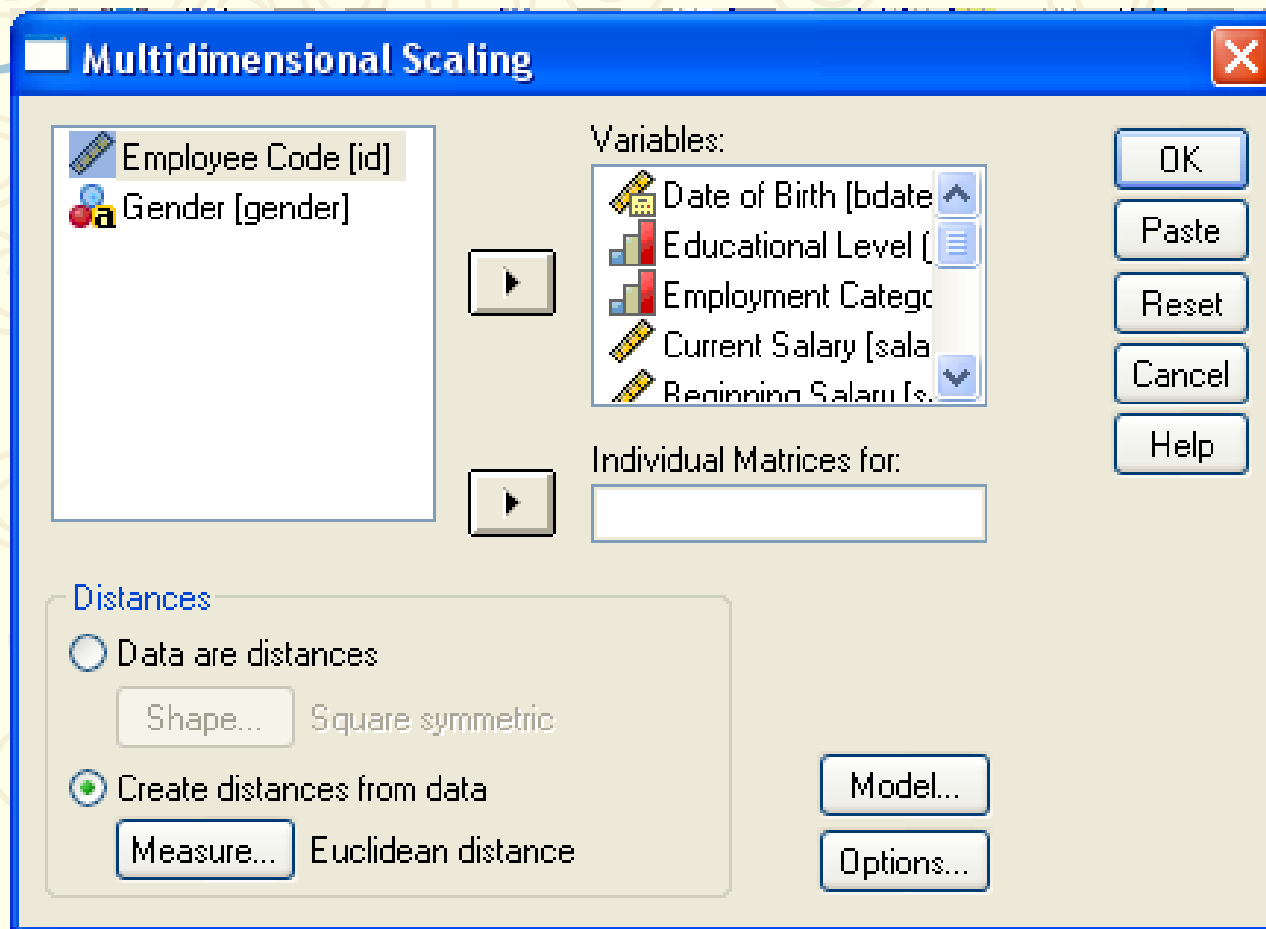
- Bir kategorinin üyeleri arasındaki benzerliğin/farklılığın altında yatan boyutlar araştırılmak istendiğinde kullanılır.
 - Sadece tek boyut varsa korelasyon kullanılır.
- Kişisel tercihler, tutumlar, eğilimler, inançlar ve bekleyişler gibi davranışsal verilerin analizinde kullanılabilir.

Boyut Sayısını Azaltmak: MDS

- Hem metrik hem de metrik olmayan deęişkenlere uygulanabilir.
 - Tam metrik MDS-aralıklı ve oransal ölçek
 - Yarı metrik ve metrik olmayan MDS-sıralı ölçek ve nominal ölçek



Boyut Sayısını Azaltmak: MDS

- Analizde göz önünde tuttuğumuz her değişken bir "boyut" sayılabilir.
- İki-üç boyuttan fazlasını uzaysal olarak görüntülemek ("visualize" etmek) kolay değil.
- Değişkenler arasındaki uzaklıktan yola çıkılarak bu boyut sayısı azaltılabilir.
- Değişkenler arasındaki uzaklıklar metrik olarak ölçülebilir




Uzaklık Matrisi

Raw (unscaled) Data for Subject 1



	1	2	3	4	5	
1	,000					
2	26,049	,000				
3	30,879	21,388	,000			
4	28,433	17,888	14,399	,000		
5	30,586	18,612	15,193	10,632	,000	
6	31,523	29,942	30,669	29,400	31,007	
7	41,247	34,377	29,753	32,187	30,018	
8	32,383	32,690	32,869	33,338	33,052	



2 Boyut için Stress istatistiğinin 0,001'den küçük olduğu değere kadar 4 iterasyon yapılmıştır.

Young's S-stress formula 1 is used.

Iteration S-stress Improvement

1	,22486	
2	,19200	,03286
3	,18706	,00495
4	,18681	,00024

Sıfıra yakın stress değerleri uygundur

Iterations stopped because
S-stress improvement is less than ,001000
Sress values are Kruskal's stress formula 1.

For matrix

Stress = ,15756 RSQ = ,86996

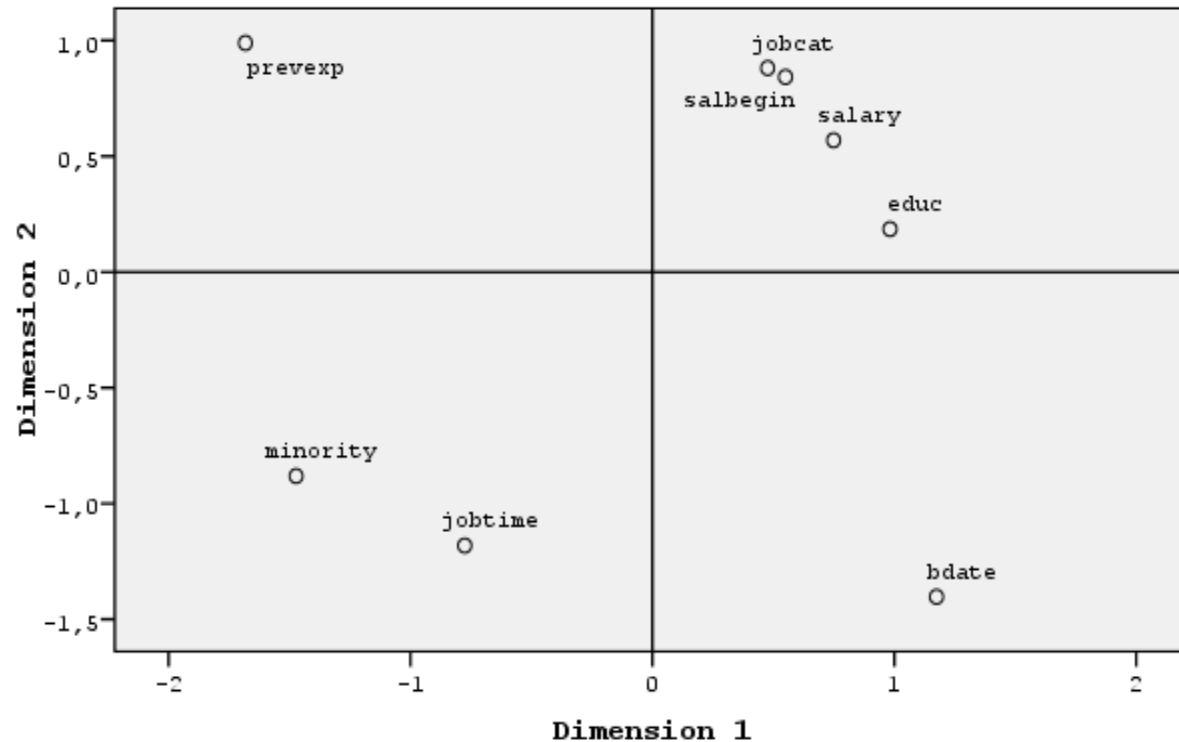
2 Boyut için stress değeri, verileri
%86,996 açıklamaktadır

Stimulus Coordinates

Stimulus Number	Stimulus Name	Dimension	
		1	2
1	bdate	1,1742	-1,4033
2	educ	,9820	,1855
3	jobcat	,4763	,8809
4	salary	,7488	,5689
5	salbegin	,5500	,8421
6	jobtime	-,7758	-1,1818
7	prevexp	-1,6824	,9890
8	minority	-1,4732	-,8814

Derived Stimulus Configuration

Euclidean distance model



Kümeleme Analizi

- Birey ya da nesnelerin anlamlı alt gruplarını oluşturmak için kullanılır. Ayırma analizinden farklı olarak gruplar önceden tanımlanmaz.
 - Değişken değil case sayısını azaltma
 - Değişkenler temelinde analize alınan "case"ler birbirlerine bir ya da daha fazla boyutta benzerler
 - Benzerliklerden yola çıkarak kümeler oluşturmak mümkün
- Pazar bölümlendirme

Kümeleme Analizi

- Değişkenlerin ölçekleri farklılık gösterebilir.
 - Farklılık varsa Z skorlarına göre standardize edilmelidir.
- Hiyerarşik ve hiyerarşik olmayan kümeleme yöntemleri bir arada kullanılabilir.
- Hiyerarşik olmayan kümeleme yönteminde küme sayısı dışarıdan girilir.

K-Means Cluster Analysis

Variables:

Yasiniz [S8]
Karsilikli baglilik [k.a]
Ortak girisim [ortakç]
Paylasilan birikim [p]

Label Cases by:

Number of Clusters: 3

Method
 Iterate and classify
 Classify only

Cluster Centers

Read initial:
 Open dataset
 External data file

Write final:
 New dataset
 Data file

Buttons: OK, Paste, Reset, Cancel, Help, Iterate..., Save..., Options...

Iteration History

Iteration	Change in Cluster Centers		
	1	2	3
1	7,196	3,293	6,530
2	,833	,206	2,818
3	,403	,126	,906
4	,000	,000	,000

- a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is ,000. The current iteration is 4. The minimum distance between initial centers is 22,076.

Cluster Membership

Case Number	Cluster	Distance
1	1	5,414
2	1	3,340
3	2	2,516
4	1	1,517
5	1	4,369
6	2	1,688
7	2	3,077
8	.	.
9	2	3,392
10	.	.
11	1	1,710
12	1	6,265
13	.	.
14	1	1,661
15	2	5,420
16	.	.
17	2	4,066
18	.	.
19	1	3,540
20	1	3,946
21	.	.
22	1	5,901
23	1	1,091
24	3	10,136
25	2	1,044
26	1	1,884
27	1	3,264
28	2	2,968
29	3	6,090
30	3	3,221
31	3	3,368
32	3	4,240
33	1	2,119
34	3	4,276
35	1	3,344

Final Cluster Centers

	Cluster		
	1	2	3
Yasiniz	22	33	48
Karsilikli baglilik	,21765	-,17870	-,09529
Ortak girisim	,15828	-,13944	-,09369
Paylasilan birikim	,02954	-,03121	,05515

Number of Cases in each Cluster

Cluster	1	153,000
	2	104,000
	3	48,000
Valid		305,000
Missing		52,000

ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Yasiniz	13254,303	2	13,775	302	962,233	,000
Karsilikli baglilik	5,334	2	,941	302	5,669	,004
Ortak girisim	3,094	2	,971	302	3,186	,043
Pay lasilan birikim	,165	2	1,034	302	,160	,852

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Conjoint Analizi

- Ürün tasarımı çalışmalarında kullanılır. Tüketicilerin hangi ürün özelliklerine önem verdiği ve hangilerini tercih edeceğini belirleme.
 - Daha çok yeni ya da gözden geçirilen bir ürün ya da hizmetin niteliklerini belirlemek, fiyatların oluşturulmasına yardımcı olmak, satış ya da kullanım düzeyini tahmin etmek ve yeni bir ürün önermek amacıyla kullanılır.
- CONJOINT ; CONsider JOINTly ; karar verirken ürün/hizmet özelliklerini bir arada dikkate almak
- Conjoint özelliklerin eş anlı olarak değerlendirilmesine dayanır.
 - Trade-off örn. bir özelliğe daha fazla sahip olmak için diğerinden ne kadar vaz geçmek gerek?

Conjoint Analizi

- Çok boyutlu ölçekleme gibi, conjoint analizi cevaplayıcıların subjektif değerlendirmelerine dayanır. MDS'de odaklanılan ürün ya da markalarken, conjoint özellik kombinasyonuna odaklanır.

$$Y = X_1 + X_2 + \dots + X_n$$

metrik, metrik olmayan
metrik olmayan

Conjoint Analizi

- Ürün ya da hizmet özellikleri ile birlikte tanımlanır
- Katılımcılara bu özelliklerin kombinasyonunu içeren ürünler, prototip ya da resimler gösterilir ve kendilerinden bu ürünleri tercihlerine göre sıralamaları istenir.
- Özellik kombinasyonlarının sayısı arttıkça olası profil sayısı da üssel olarak artış gösterir.

Conjoint Analizi

Alternatif	Konfor	Fiyat	Süre
1	dar koltuk	\$225	2 saat
2	dar koltuk	\$225	5 saat
3	dar koltuk	\$800	2 saat
4	dar koltuk	\$800	5 saat
5	geniş koltuk	\$225	2 saat
6	geniş koltuk	\$225	5 saat
7	geniş koltuk	\$800	2 saat
8	geniş koltuk	\$800	5 saat

En az tercih edilen



En çok tercih edilen

Correspondence Analysis

■ Amaç:

- Verilerin görsel sunumu
- İki sayısal değişken arasındaki ilişkiyi tanımlamak ve aynı zamanda her değişkenin kategorileri arasındaki ilişkileri belirlemek.

■ Veri tipi: Nonmetrik. Özellikle kategorik verilerin analizinde uygundur.

■ Bir tablodaki karşılıklı etkilerin grafik özetini sunar. Her değişken için, düzleme yansıtılan kategoriler arasındaki mesafe yakınlık ilişkisini ifade eder.

■ Benzer bir çok grafik gibi, bu da algı haritası olarak da bilinir.

Correspondence Analysis

- Correspondence analiz grafikleri, noktaların orijine göre incelenerek yorumuna dayanır.
 - Benzer yönlerdeki noktalar pozitif ilişkilidir.
 - Ters yönlerdeki noktalar negatif ilişkilidir.
 - Orijinden en uzak noktalar en güçlü ilişkiyi gösteren noktalardır.

CNN TURK

◆ ◆ Ekonomi prog.

◆ **NTV**

◆ **TV8**

Açık oturum

KANAL E /CNBC-E

◆ Haberler
◆ Hava durumu

ATV

Yerli diziler

◆ Yabancı diziler

◆ **SHOW**

◆ Magazin prog.

KANAL7 ◆

◆ **TGRT**

Spor prog.

◆ **STAR**

◆ **TRT 1**

◆ **TRT 3**

◆ **TRT 2**

◆ Belgesel

◆ **STV (Samanyolu TV)**

◆ **KANAL 6**

◆ Yarışma prog.

◆ **KANAL D**

◆ Talk show prog.

◆ Çocuk prog.

Yapısal Eşitlik Modelleme (Structural Equation Modelling)

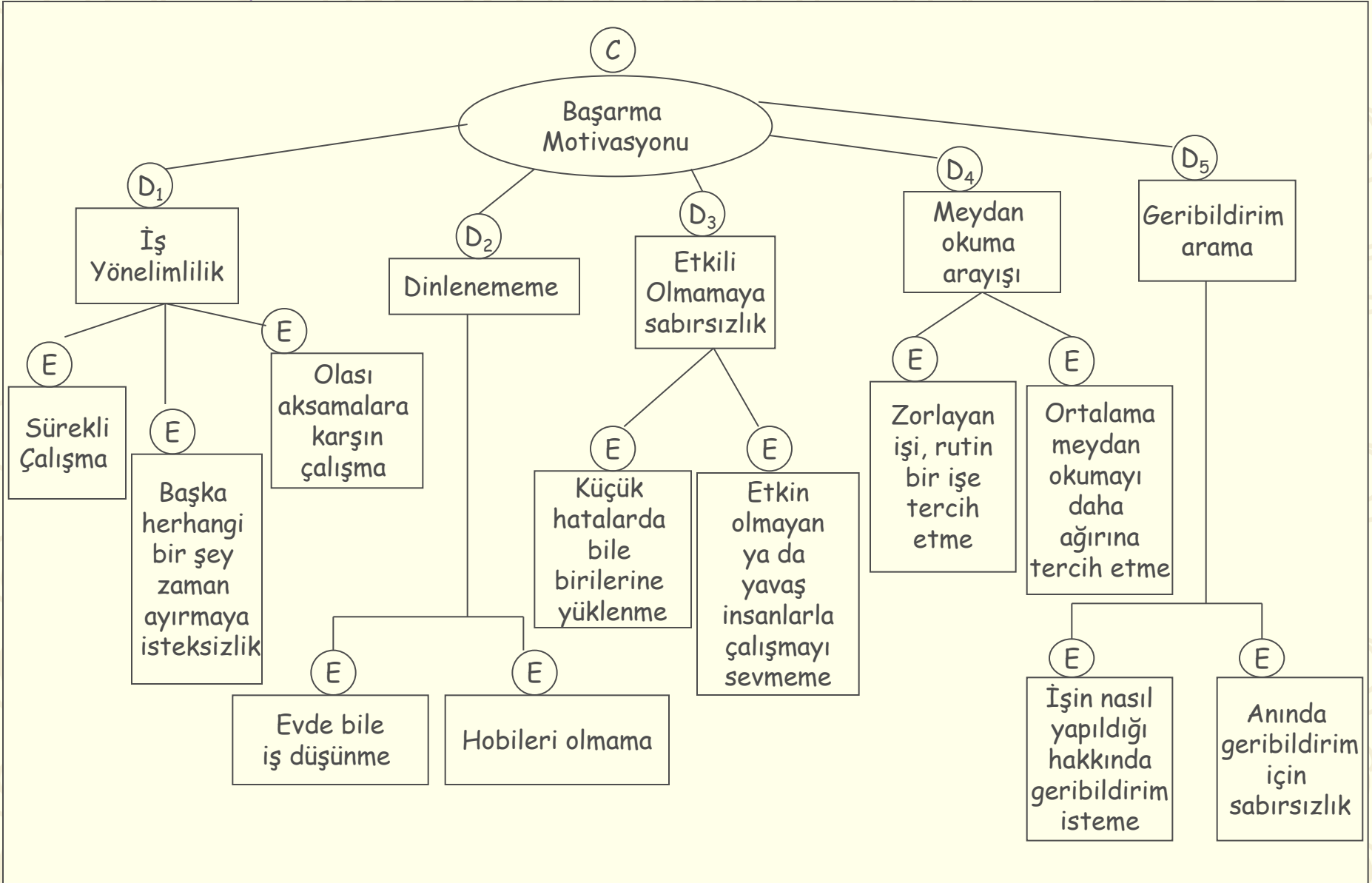
- Her bir bağımlı değişken seti için ilişkileri ayırdetmek için kullanılır. İki temel bileşeni vardır; bağımsız değişkenleri bağımlı değişkenler ile ilişkilendiren yapısal model ve tek bir bağımsız ya da bağımlı değişken için araştırmacının bir çok değişkeni kullanmasına olanak sağlayan yapısal model.

$$\begin{array}{l} Y_1 \\ \text{metrik} \end{array} = \begin{array}{l} X_{11} + X_{12} + \dots + X_{1n} \\ \text{metrik} + \text{metrik olmayan} \end{array}$$
$$\begin{array}{l} Y_2 \\ \text{metrik} \end{array} = \begin{array}{l} X_{21} + X_{22} + \dots + X_{2n} \\ \text{metrik} + \text{metrik olmayan} \end{array}$$
$$\begin{array}{l} Y_m \\ \text{metrik} \end{array} = \begin{array}{l} X_{m1} + X_{m2} + \dots + X_{mn} \\ \text{metrik} + \text{metrik olmayan} \end{array}$$

Latent Değişkenler

- SEM'in gerçek gücü *gizil (latent) değişken modellemeyen* gelmektedir.
- Varsayım; gözlenen ve gözlenemeyen (gizil-latent) değişkenler arasındaki nedensel ilişkiler test edilirken gizil değişkenler gözlenen değişkenler aracılığıyla ölçülebilir
- Bu tip değişkenler hiç bir zaman doğrudan ölçülemeyen değişkenlerdir.
 - Etkileri ölçülür
 - Tutumlar
 - Zeka (Intelligence)
 - Tatmin
 - Tercih
 - Motivasyon

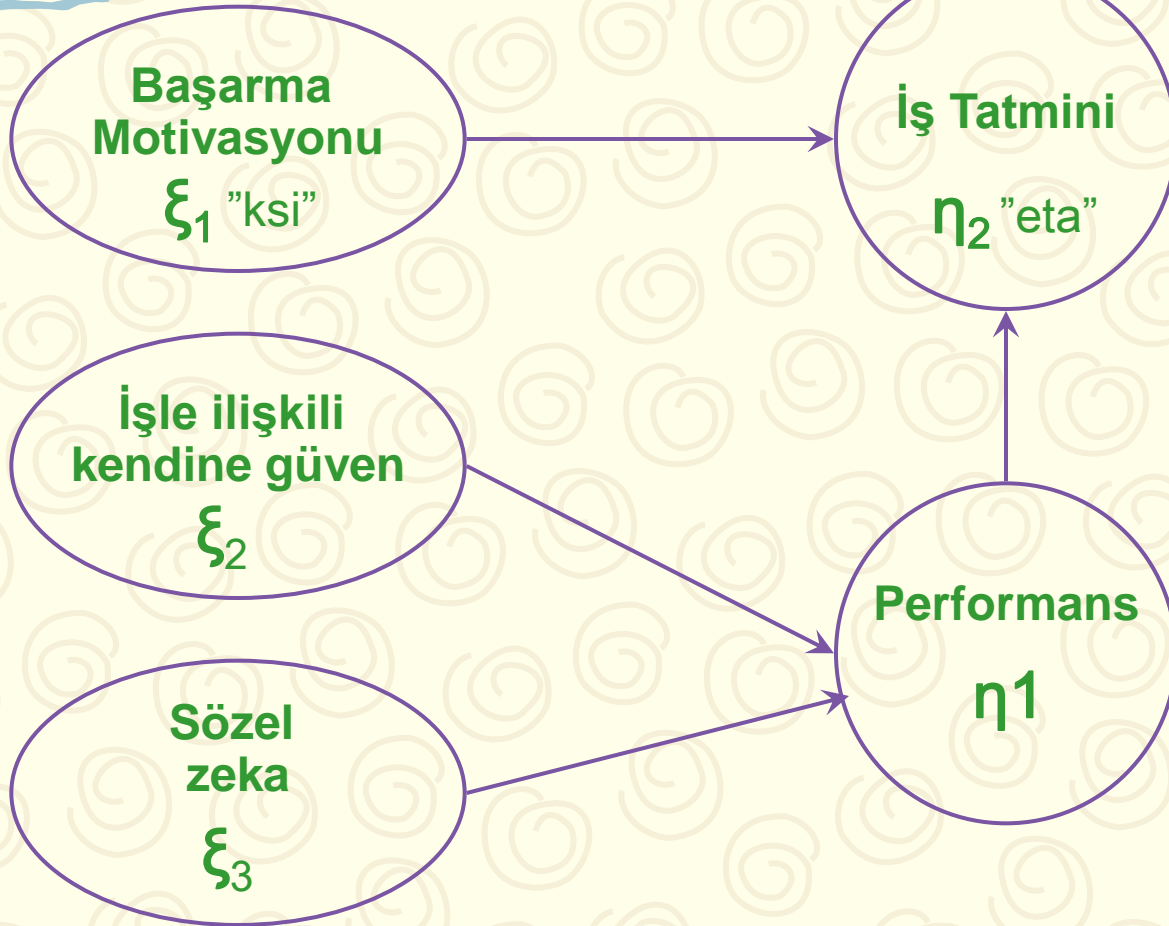
Boyutlar (D), Elemanlar (E) ve Kavramlar (C)



Örnek

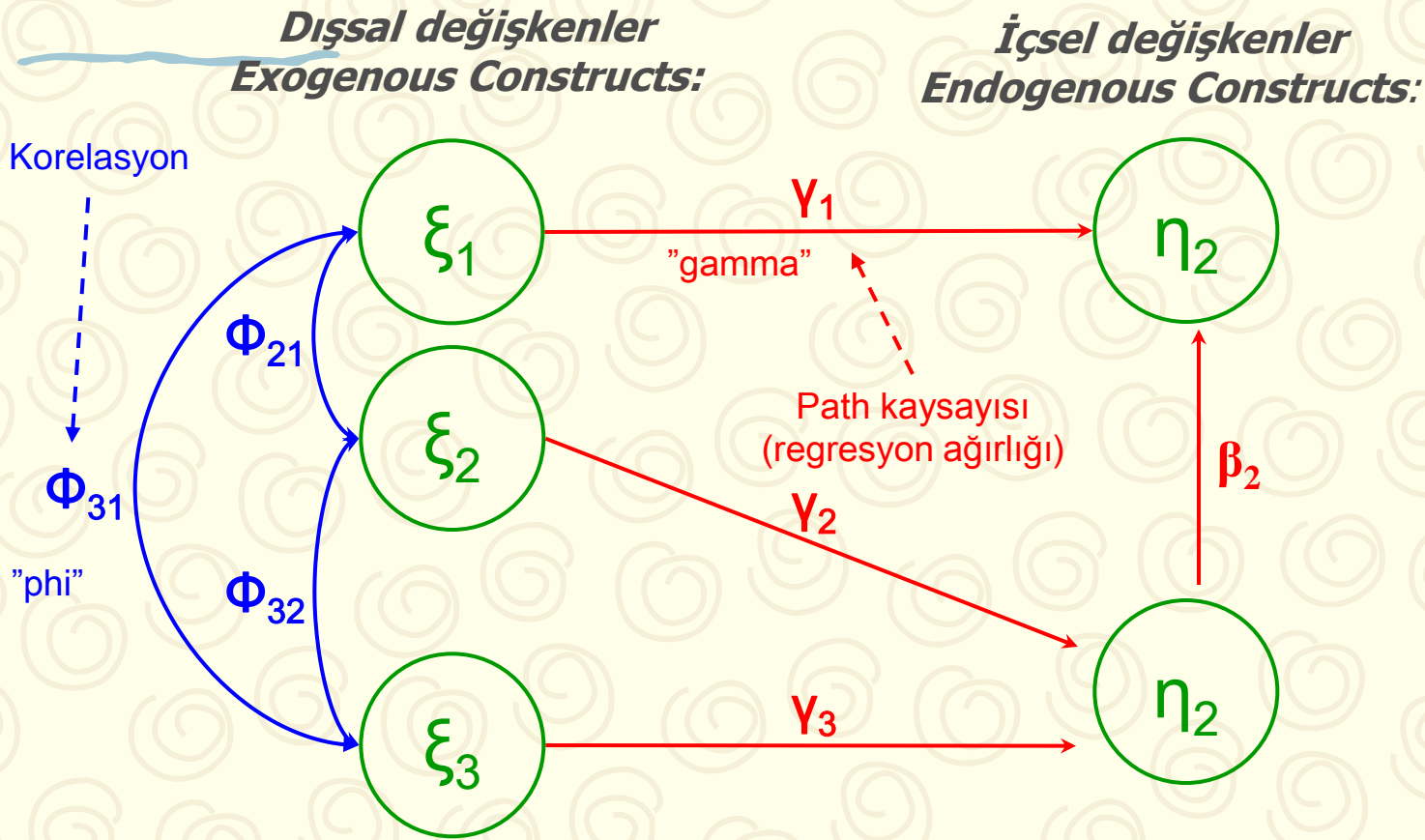
**Girdiler (Bağımsız –
dışsal değişkenler)**

**Çıktılar (Bağımlı-
içsel değişkenler)**



Rota diyagramının (path model) kurulması

Gizil değişkenler ve nedensel ilişkiler



- Rota katsayıları için yüklerinin hesaplanması sonucunda her birinin t değerleri 2'den daha büyük olmalı
 - Anlamlı sonuçlar, değişkenlerin istatistiki olarak belirlenen yapılar ile ilişkili olduğunu gösterir.
- Gizil yapılar arasındaki korelasyon incelenir.
- Standart hatalar, parametrelerin değerlerinin doğru bir şekilde nasıl tahmin edildiğini gösterir. Standart hata ne kadar küçükse tahminler de o derece uygundur.