



Yol Analizi



Model Tanımlama

Dr. Kubra Atalay Kabasakal
Bahar 2023

Bir Araştırma Senaryosu: Hastalık Faktörleri

- Veri setini okuma

```
library(readr)
veri <- read_table("illness.dat", col_names = FALSE)
colnames(veri) <- c("form", "stres", "hastalik", "egzersiz", "dayaniklilik")
```

- Yol modelini kurma

```
library(lavaan)
yol_model <- 'stres      ~ egzersiz + dayaniklilik
              hastalik  ~ egzersiz + dayaniklilik + form + stres
              form      ~ egzersiz + dayaniklilik
egzersiz ~~ dayaniklilik'
yol_fit <- sem(yol_model, veri)
```

Yol Analizi Modeli

- Veri modellemedeki düşünce **gözlenen kovaryans matrisinin bir grup değişken arasındaki varsayılan ilişki tarafından üretilip üretilmeyeceğini** test etmektir.
- Varsayılan model belli bir varyans ve kovaryans deseni gerektirir ki bu varyans ve kovaryans deseni **üretilmiş varyans ve kovaryans** matrisi (reproduced variance and covariance matrix) olarak adlandırılır.
- Bu matris çoğunlukla **üretilmiş kovaryans matrisi** (reproduced covariance matrix) olarak kısaltılır.

Yol Analizi Modeli

- Gözlenen kovaryans matrisi ve üretilmiş kovaryans matrisi arasındaki fark **artık kovaryans matrisini** (residual covariance matrix) oluşturur.
- Eğer **artık kovaryans matrisinin** bütün elemanları **sıfır**sa, varsayılan model veriyle **tamamen eşleşmiş** demektir (mükemmel model-veri uyumu)
- Eğer artık kovaryans matrisinin bütün elemanları **sıfır değilse**, varsayılan **model ve veri arasında bir takım uyumsuzluklar vardır.**

Yol Analizi Modeli

- Yol analizi modelinde **sıfır hipotezi**, model tarafından üretilen **kovaryans matrisinin gerçek veriyle (örneğin, gözlenen kovaryans matrisiyle)** tamamen eşleştiğidir.

$$\Sigma = \hat{\Sigma}$$

Burada

- Σ (sigma): gözlenen değişkenlerin evren kovaryans matrisi
- $\hat{\Sigma}$: model tarafından üretilen kovaryans matrisi
- Yol analizi modelinde, sıfır hipotezinin reddedilmesi istenmez! Bu nedenle, genel model uyumunu test etmek için kullanılan **olabilirlik oranı (ki-kare)** testinin **p-değerinin yüksek olması istenir**.

Yol Analizi Modeli

- **Artık kovaryans matrisi** model tarafından açıklanmayan varyans ve kovaryansları içeren matristir. Aşağıdaki eşitlikle hesaplanır:

$$S - \hat{\Sigma}$$

S örneklemden elde edilen gözlenen kovaryans matrisidir.

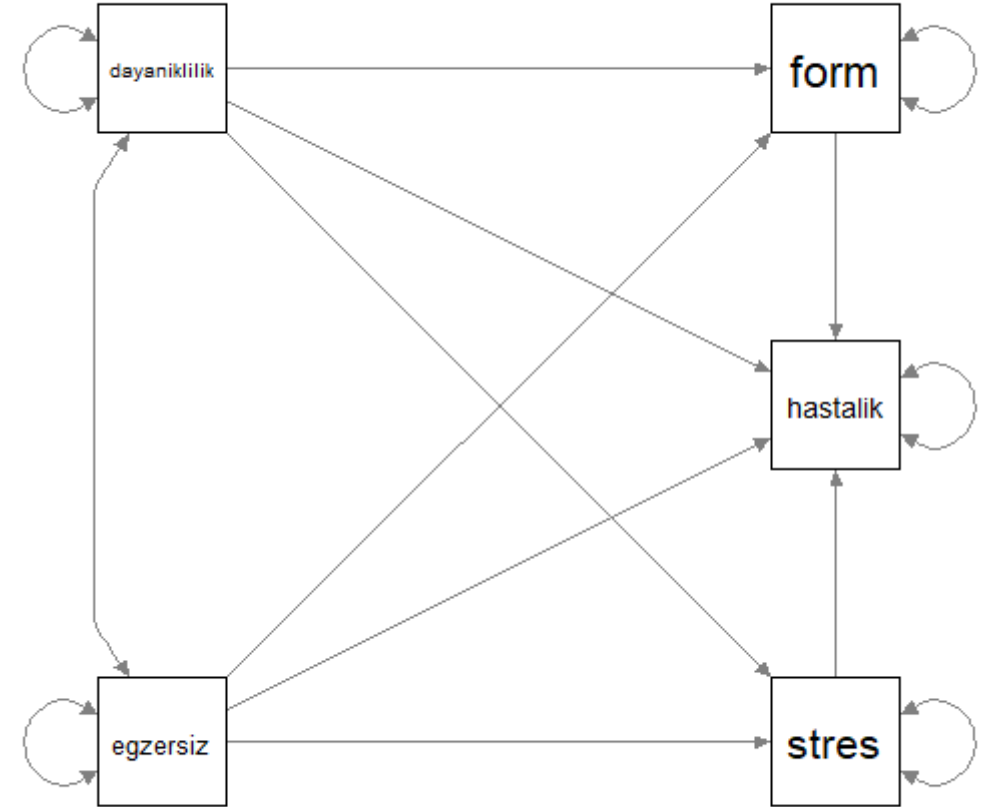
- Gözlenen değişkenler arasındaki bütün varyans ve kovaryanslar model tarafından açıklandığı zaman artık **kovaryans matrisinin bütün öğeleri sıfır olacaktır.**
- Aksi halde, artık kovaryans matrisinin sıfırdan farklı olacaktır.

Bir Araştırma Senaryosu:

Hastalık Faktörleri (Yol Şeması)

```
library(semPlot)

semPaths(yol_fit,
  rotation=2,
  sizeMan = 10,
  edge.label.cex = 1.15,
  style = "ram")
```



Evren Kovaryans Matrisi

- hastalık faktörleri örneği için gözlenen değişkenlerin evren kovaryans matrisi Σ aşağıdaki gibidir:

$$\begin{bmatrix} VAR_{y_1} & & & & \\ COV_{y_2,y_1} & VAR_{y_2} & & & \\ COV_{y_3,y_1} & COV_{y_3,y_2} & VAR_{y_3} & & \\ COV_{x_1,y_1} & COV_{x_1,y_2} & COV_{x_1,y_3} & VAR_{x_1} & \\ COV_{x_2,y_1} & COV_{x_2,y_2} & COV_{x_2,y_3} & COV_{x_2,x_1} & VAR_{x_2} \end{bmatrix}$$

- Σ kare ve simetrik bir matristir ve matriste çoğunlukla alt üçgen ve köşegen öğeleri yazılır.

Model gözlemleri:

- bir yol modelindeki gözlenen değişkenler arasındaki varyans ve kovaryanslardır.
- Model gözlemlerinin sayısı v gözlenen değişkenlerin sayısı olmak üzere $v(v + 1)/2$ 'ye eşittir.
- hastalık faktörleri örneğine göre, $v = 5$ ve $v(v + 1)/2 = 15$

$$\begin{bmatrix} VAR_{y_1} & & & & \\ COV_{y_2,y_1} & VAR_{y_2} & & & \\ COV_{y_3,y_1} & COV_{y_3,y_2} & VAR_{y_3} & & \\ COV_{x_1,y_1} & COV_{x_1,y_2} & COV_{x_1,y_3} & VAR_{x_1} & \\ COV_{x_2,y_1} & COV_{x_2,y_2} & COV_{x_2,y_3} & COV_{x_2,x_1} & VAR_{x_2} \end{bmatrix}$$

Model Parametreleri:

Modelden kestirilen parametre sayısı

- ya gözlenen ya da gözlenmeyen **dışsal değişkenlerin varyanslarının ve kovaryanslarının sayısı** ve
- **gözlenen değişkenlerden içsel değişkenlere olan doğrudan etkilerin sayısı** toplamıdır.

Model Parametreleri:

- Hastalık faktörleri örneğinde, kestirilecek model parametreleri:
- Dışsal değişkenlerin varyansları: 2
- Dışsal değişkenlerin kovaryansı: 1
- Bozuklukların varyansları: 3
- Doğrudan etkiler: 8

$$2 + 1 + 3 + 8 = 14$$

İçsel değişkenlerin varyansları ve kovaryansları model parametreleri olarak ele alınmaz.

Bozukluk, modelde nedeni bilinmediğinden, gözlenmeyen dışsal değişken olarak ele alınır.

Model Parametreleri:

- Model parametresi araştırmacının tanımlamasına bağlı olarak **serbest** (free), **sabit** (fixed) veya **sınırlandırılmış** (constrained) olabilir.
- **Serbest parametre** (free parameter) örneklem verisinden bilgisayar yazılımı tarafından kestirilen parametredir.
- **Sabit** parametre (fixed parameter) bir sabite eşit olarak belirlenen parametredir; yazılım bu sabiti veriye bağlı olmaksızın parametrenin kestirimi olarak kabul eder.
- **Sınırlandırılmış** parametre (constrained parameter) yazılım tarafından belli sınırlılıklar içerisinde kestirilir ancak bir sabite eşit olmak üzere sabitlenmez.

Serbestlik Derecesi:

- **Model serbestlik derecesi (sd)** model gözlemlerinin sayısı ve modelden kestirilecek parametre sayısı arasındaki farka eşittir:
- Hastalık faktörleri örneğinde,

$$sd = 15 - 14 = 1$$

- **sd<0** model tanımlanamaz.
- **sd=0** model ancak tanımlanır (just identification) ve kuramsal olarak her parametrenin tek bir çözümü vardır. Model veriye mükemmel uyum gösterir.
- **sd>0** model aşırı tanımlanmış (over identification) olur. Aşırı tanımlanan modellerde kuramsal olarak her bir parametrenin birden fazla çözümü vardır.

Model Tanımlanması

(Model Identification) İlkeleri

- Bir yol modelinin değerlendirilmesi sırasında karşılaşılabilecek olası problemlerden biri **modelin tanımlanmasıdır**.
- Modeldeki **her bir parametre için** kuramsal olarak **tek bir kestirim** üretilebiliyorsa, model tanımlanır denir. Aksi halde, model tanımlanamaz.
- **Kuramsal olarak ifadesi tanımlanmanın verinin değil, modelin bir özelliği** olduğunu vurgulamak için kullanılmıştır.

Model Tanımlanması

(Model Identification) İlkeleri

- Örneğin, **eğer bir model tanımlanmazsa, örneklem büyüklüğüne (100, 1000, vb.) bağlı olmaksızın** tanımlanamaz olarak kalır.
- Bu nedenle, **tanımlanmayan modellerin yeniden tanımlanması** gerekir. Aksi halde, **analizler çözüm** üretmez.
- Herhangi bir yapısal eşitlik modeli için tanımlanmanın gerektirdiği iki koşul vardır:
 - Her **gözlenmeyen (gizil) değişken bir ölçeğe atanmalıdır**, bozukluk yol analizindeki tek gizil değişkendir.
 - **En az serbest model parametreleri kadar gözlem olmalıdır ($sd \geq 0$)**

Model Tanımlanması

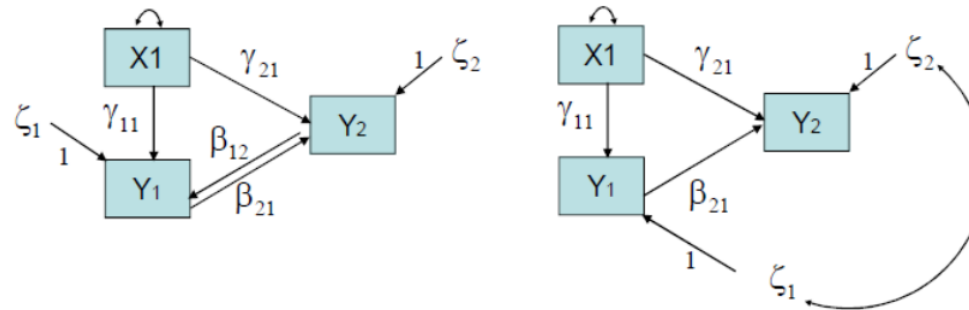
Gizil Değişkenlere Ölçek Atanması

- Gizil değişkenler **yapay bir ölçektedirler**. Yazılımın gizil değişkenleri içeren etkilerin kestirimlerini hesaplayabilmesi için **gizil değişkenlere bir ölçek atanması gerekir**.
- **Unit Loading Identification (ULI) constraint:** Artık yol katsayısı (bozukluğun doğrudan etkisini ifade eden yol katsayısı) 1.0'e sabitlenir, böylece bozukluk ilgili içsel değişkenin açıklanmayan varyansı ile ilişkili bir ölçeğe atanmış olur.
 - Regresyon ve yol analizi modellerinde, **ölçekler bozukluklara genellikle ULI** aracılığıyla atanırlar.
- **Unit Variance Identification (UVI) constraint:** Gizil içsel değişkenlerin (yol modelinde bozuklukların) varyansı 1.0'e eşitlenir böylece gizil değişkenin varyansı standartlaştırılmış ölçekte olur.

Model Tanımlanması

Yetersiz Tanımlanma (Underidentification)

- Yetersiz tanımlanan bir modelde **serbest model parametrelerinin** sayısı gözlem sayısından büyüktür, **diğer bir ifadeyle modelin serbestlik derecesi sıfırdan küçüktür ($sd < 0$)**.
- Yetersiz tanımlanan bir model test edilemez ve yeniden tanımlanması gerekir. Bu durumda en az bir parametrenin çözümü yoktur.
- Örnek:



Model Tanımlanması

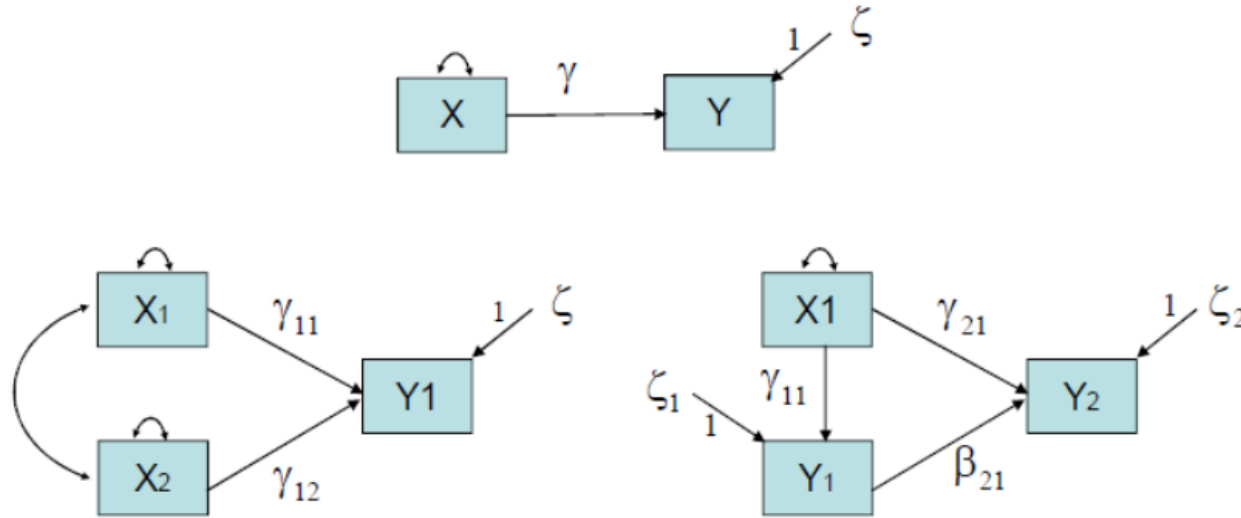
Yetersiz Tanımlanma (Underidentification)

- Gözlem eksikliğinin tanımlanmamaya nasıl yol açtığının bir örneği
 - $a + b = 6$
- Verilen eşitlik bir model olarak ele alınırsa, **6 gözlem**, **a ve b de parametrelerdir**.
- Eşitlikte **parametre sayısı (2)**, **gözlem sayısından (1)** daha fazla olduğundan, her bir parametre için tek bir çözüm bulmak imkansızdır.
- **a ve b** parametreleri için eşitliği sağlayacak sonsuz sayıda çözüm vardır:
- Örneğin, $(a = 4, b = 2)$, $(a = 8, b = -2)$ vb.
- **Yetersiz tanımlanan bir yol modeli için** program **her bir parametreye ait tek bir kestirim üretmeye** çalışırken de benzer bir durum söz konusudur.

Model Tanımlanması

Ancak Tanımlanma (Just Identification)

- Ancak tanımlanan bir modelde **parametre sayısı ve gözlem sayısı** birbirine eşittir, diğer bir ifadeyle modelin serbestlik derecesi sıfıra eşittir ($sd = 0$) ve model tanımlanır; kuramsal olarak her parametrenin tek bir çözümü vardır.
- Örnek:



Model Tanımlanması

Ancak Tanımlanma (Just Identification)

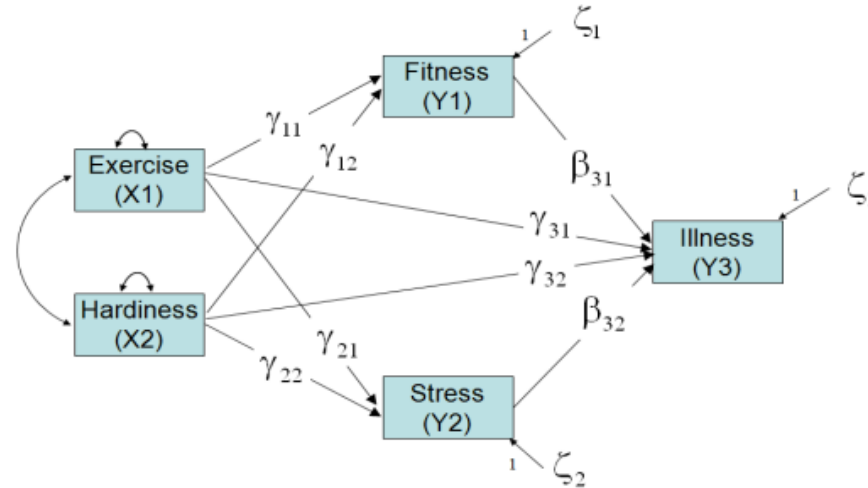
Ancak tanımlanmanın bir örneği aşağıda verilmiştir:

- $a + b = 6$
- $2a + b = 10$
- Verilen eşitlikler bir model olarak ele alınırsa, **6** ve **10 gözlemler**, **a** ve **b de parametrelerdir**.
- Eşitlikte **parametre sayısı (2)**, **gözlem sayısına (2)** eşit olduğundan, her bir parametre için tek bir çözümü bulunmaktadır.
 - $(a = 4, b = 2)$
 - Parametre kestirimleri verildiğinde gözlemler mükemmel bir şekilde üretilir.
- Ancak tanımlanan bir yol modeli için program her bir parametreye ait tek bir kestirim üretmekle kalmayıp **model veriye mükemmel uyum** sağlayacaktır.

Model Tanımlanması

Aşırı Tanımlanma (Overidentification)

- Aşırı tanımlanan bir modelde **parametre sayısı, gözlem sayısından küçüktür**, diğer bir ifadeyle modelin serbestlik derecesi sıfırdan büyüktür **$sd > 0$** , kuramsal olarak her parametrenin olası bir çok çözümü vardır.



- Aşırı tanımlanan bir model için, üretilen kovaryans matrisi örneklem kovaryans matrisini **mükemmel bir şekilde üretmeyecektir**.

Model Tanımlanması

Aşırı Tanımlanma (Overidentification)

Aşırı tanımlanmanın bir örneği aşağıda verilmiştir:

- $a + b = 6$
- $2a + b = 10$
- $3a + b = 12$
- Verilen eşitlikler bir model olarak ele alınırsa, **6, 10 ve 12** gözlemler, **a ve b de parametrelerdir.**
- Üç eşitliği sağlayacak **a ve b değerleri bulunmayabilir.**
- Örneğin, $(a = 4, b = 2)$ değerleri sadece ilk iki eşitliği sağlar.

Model Tanımlanması

Aşırı Tanımlanma (Overidentification)

- Aşırı tanımlama durumunda, her bir parametre için tek bir çözüm aşağıdaki şekilde üretilir:
 - Gözlemler ve üretilen gözlemler arasındaki farkın karesinin toplamının mümkün olduğunca küçük olacağı pozitif parametre değerleri bulunur.
 - Örneğin, $(a = 3, b = 3,3)$ sadece en küçük toplam kareler farkını sağlamakla kalmaz tek bir çözüm üretir.

bitti