📉 Çoklu Regresyon

📈 Bir Örnek

Dr. Kubra Atalay Kabasakal

Bahar 2023

Çoklu Regresyon

- Çok değişkenli analiz, bir çalışmadaki bireylerden veya nesnelerden elde edilen çoklu ölçümlerin aynı anda analizidir. Dolayısıyla **ikiden fazla değişkenin** aynı anda analizi çok değişkenli analiz olarak düşünülebilir.
- Çok değişkenli analiz yöntemlerinin çoğu **tek veya iki değişkenli** analiz yöntemlerinin uzantısıdır.
- Örneğin, tek bir bağımsız (yordayıcı) değişkenle gerçekleştirilen basit regresyon, **birden fazla bağımsız (yordayıcı) değişkenle** gerçekleştirilen **çoklu regresyona** genişletilmiştir.
- Benzer şekilde, tek bir bağımlı (yordanan) değişkenin bulunduğu varyans analizi, birden fazla bağımlı (yordanan) değişkenin bulunduğu çok değişkenli varyans analizine genişletilmiştir.

Çoklu Regresyon

- Çok değişkenli istatistiksel analiz yöntemleri aşağıdaki şekilde sınıflanabilir:
- Eş zamanlı regresyon eşitlikleri
- Gizil değişken modellemesi
- Sınıflamalar
- Tekrarlı ölçümler
  - Not: Detaylar için bkz. Tabachnick & Fidell (2007), s. 29- 30).

Çoklu Regresyon

- Çoklu regresyon, basit regresyonun tek bir bağımlı değişkenin **iki veya daha fazla yordayıcısına** izin veren uzantısıdır. Diğer bir ifadeyle, çoklu regresyon tek bir bağımlı değişken ile **iki veya daha fazla bağımsız (yordayıcı)** değişken arasındaki ilişkinin analiz edilmesi için kullanılan istatistiksel bir yöntemdir.
- Çoklu regresyonun amacı değerleri bilinen **bağımsız değişkenleri kullanarak bağımlı değişkenin değerini** yordamaktır.
- Regresyon yöntemiyle **bağımsız değişkenlerden en fazla yordamayı sağlamak üzere her bağımsız değişken** ağırlıklandırılır.

Çoklu Regresyon

- Ağırlıklar bağımsız değişkenin **yordamaya bağıl katkısını** ifade eder ve her bir değişkenin yordamadaki etkisine ilişkin yorumlamayı kolaylaştırır.

- Çoklu regresyon, hem bağımlı değişken hem de bağımsız değişkenler **en az eşit aralıklı ölçek düzeyinde** ölçüldüğünde kullanılmalıdır.

- Ancak bağımsız değişkenler **sınıflama veya sıralama ölçeğinde** ölçüldüğünde ilgili değişkenler **belli koşullar** altında analize dahil edilebilir.

Çoklu Regresyon

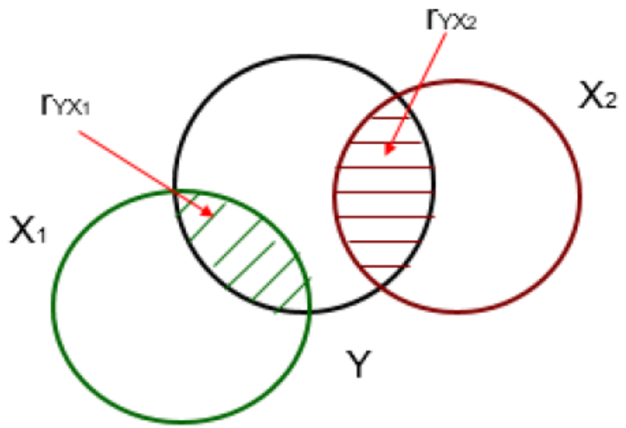
- Çoklu regresyon her bir bağımsız değişkendeki değişikliklerin **bağımlı değişkendeki değişikliklerle ne ölçüde** ilişkili olduğunu kestirir.

- Ancak **bağımsız değişkenler arasındaki korelasyon** yordama sürecini zorlaştırır.

Çoklu Regresyon

Örneğin, X_1 ve Y arasındaki korelasyon katsayısı **0.40**, X_2 ve Y arasındaki korelasyon katsayısı **0.60**, X_1 ve X_2 arasındaki korelasyon katsayısı sıfır ise, Y'nin varyansının iki değişken tarafından açıklanan toplam oranı iki değişkenin Y ile korelasyonlarının kareleri toplamından elde edilebilir:

$$0.40^2 + 0.50^2 = 0.16 + 0.25 = 0.41$$

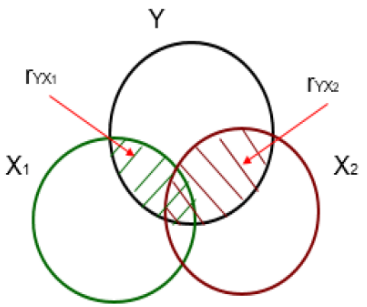


Çoklu Regresyon

- Ancak, uygulamada çoğunlukla X_1 ve X_2 birlikte değişim gösterirler ve iki değişkenin Y ile korelasyonlarının kareleri toplamı çok yüksek bir oran verir.

- Bunun nedeni, iki bağımsız değişkenin aralarındaki korelasyondan dolayı her bir bağımsız değişken tarafından açıklanan Y varyansının bir kısmının üst üste gelmesidir.

- Çoklu regresyonun en önemli özelliği modele eklenen bağımsız değişkenler arasındaki ilişkileri kontrol altına almasıdır.



Çoklu Regresyon

ceteris paribus

- Modeldeki bağımsız değişkenler arasındaki ilişkilerin kontrol altına alınması, modeldeki bir değişkenin bağımlı değişken uzerindeki etkisini incelerken, **modeldeki diğer bütün değişkenlerin sabit** tutulmasıdır.
  - Örneğin, bir çalışmada kahve tüketiminin ölüm oranını nasıl etkilediği çalışılmıştır. Basta, sonuçlar daha yüksek kahve tüketiminin daha yüksek olum riskiyle ilişkili olduğunu göstermiştir. Ancak kahve içen çoğu kişi sigara da içmektedir. Araştırmacılar modellerine sigara içme alışkanlıkları için bir değişken eklediklerinde, sigara içmenin olum riskini artırırken, kahve tüketiminin ölüm riskini azalttığını bulmuşlardır.
- Bu durumda modele **bütün önemli değişkenlerin** eklenmesi gerekmektedir. Önemli değişkenlerin modelin dışında bırakılması, katsayılara ilişkin kestirimlerin **yanlı** olmasına neden olabilmektedir.

Çoklu Regresyon

- **Performans:** Öğrencilerin matematik performans düzeyleri olup eşit aralık ölçeğinde ölçülen sürekli bir değişkendir.
- **Motivasyon:** Öğrencilerin motivasyon düzeyleri olup eşit aralık ölçeğinde ölçülen sürekli bir değişkendir.
- **Kaygı:** Öğrencilerin kaygı düzeyleri olup eşit aralık ölçeğinde ölçülen sürekli bir değişkendir.
- **Güven:** Öğrencilerin matematiğe karşı güven düzeyleri olup eşit aralık ölçeğinde ölçülen sürekli bir değişkendir.

Betimsel İstatistikler

```r
library(haven)
performans <- read_sav("Performans.sav")
psych::describe(performans)[,3:4]
```

```
##             mean    sd
## Performans 18.2  7.83
## Motivasyon 39.9 10.02
## Kaygi      18.1  4.77
## Guven      21.6  7.37
```

Çoklu Regresyon

Korelasyon değerleri ve anlamlılığı

```
cor_1 <- cor.test(~ Performans + Motivasyon , data = performans)
broom::tidy(cor_1)[,c(1,3)]
```

```
## # A tibble: 1 × 2
##   estimate  p.value
##      <dbl>    <dbl>
## 1    0.824 0.000156
```

Çoklu Regresyon

Korelasyon değerleri ve anlamlılığı

```
cor_2 <- cor.test(~ Performans + Kaygi , data = performans)
broom::tidy(cor_2)[,c(1,3)]
```

```
## # A tibble: 1 × 2
##   estimate p.value
##      <dbl>   <dbl>
## 1   -0.241   0.388
```

Çoklu Regresyon

Korelasyon değerleri ve anlamlılığı

```
cor_3 <- cor.test(~ Motivasyon + Kaygi , data = performans)
broom::tidy(cor_3)[,c(1,3)]
```

```
## # A tibble: 1 × 2
##   estimate p.value
##      <dbl>   <dbl>
## 1    0.147   0.601
```

Çoklu Regresyon

- Öğrencilerin matematikteki **performans düzeylerini, motivasyon ve kaygı düzeylerinden yordamak** ile ilgilendiğimizi düşünelim.

- Bu araştırma sorusuna cevap vermek için çoklu regresyon uygun bir istatistiksel analiz yöntemdir.

$$Y_{performans_i} = b_0 + b_1 X_{motivasyon_i} + b_2 X_{kaygi_i} + e_i$$

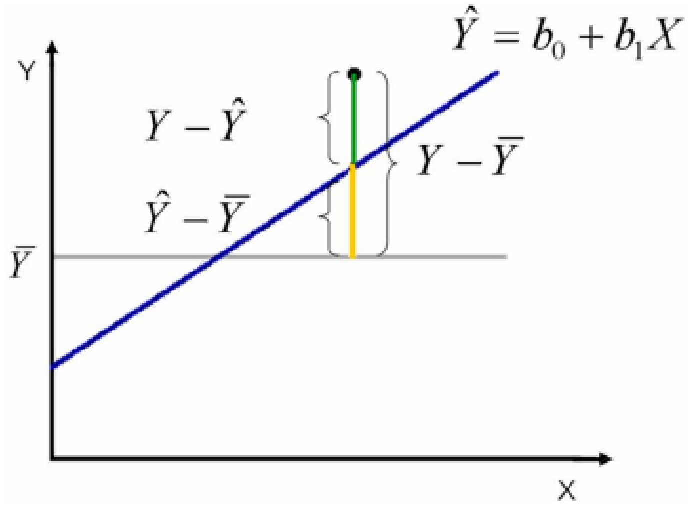
- Burada, b_1 ve b_2 motivasyon ve kaygı yordayıcıları için ağırlıklardır. Diğer bir ifadeyle regresyon katsayılarıdır veya eğimlerdir. b_0 ise kesişimdir.

Çoklu Regresyon

- Amaç hata puanlarının (artıkların) kareleri toplamının küçüleceği, diğer bir ifade ile Y ve yordanan Y' arasındaki korelasyonun büyüyeceği, b_0, b_1 ve b_2 değerleri için tek bir çözüm kümesi bulmaktır.

- Grafiğin sadeleştirilmesi için bir bağımsız değişken kullanılmıştır.

- Tek bir çözüm bulmak için kullanılan yöntem Sıradan **En Küçük Kareler Yöntemi** (Ordinary Least Squares Procedure) olarak adlandırılır.



Çoklu Regresyon

- R^2 değeri *çoklu korelasyon katsayısı* (multiple correlation coefficient) olup bağımlı değişkenin gözlenen değerleri ile bağımsız değişkenlerin *en iyi doğrusal kombinasyonu* arasındaki korelasyondur.

- **En iyi doğrusal kombinasyon**, bağımlı değişkenin bağımsız değişkenlerden yordanmasında, daha iyi bir iş yapacak regresyon katsayıları kümesi olmadığı anlamına gelir.

Çoklu Regresyon

```r
library(broom)
model <- lm(Performans ~ Motivasyon + Kaygi,
            data=performans)
sqrt(glance(model)[,1])
```

```
## # A tibble: 1 × 1
##   r.squared
##       <dbl>
## 1     0.902
```

- R değeri bağımlı değişkenin **gözlenen** ve **yordanan** değerleri arasındaki **korelasyondur**.
- Bağımlı değişkenin **yordanan değerinin** bağımlı değişkenin **gözlenen değerine** mümkün olduğunca yakın olmasını gerektiren **en küçük kareler kriterinden** dolayı bağımlı değişkenin **gözlenen ve yordanan değerleri arasındaki korelasyon eksi değerler** alamaz. Dolayısıyla çoklu korelasyon katsayısı 0 ile 1 arasında değişir.

Çoklu Regresyon

Çoklu Korelasyon

$$R_{Y_{12}} = \sqrt{\frac{r^2_{Y_1} + r^2_{Y_2} - 2r^2_{Y_1} r^2_{Y_2} r_{12}}{1 - r_{12}}}$$

$$R_{Y_{12}} = \sqrt{\frac{(0.824)^2 + (-0.241)^2 - 2*(0.824)(-0.241)(0.147)}{1 - (0.147)^2}}) = 0.902$$

```
model_s <- augment(model,data=performans)
cor(model_s[,1], model_s[,5])
```

```
##              .fitted
## Performans   0.902
```

öğrencilerin **gözlenen performans puanları** ve **yoradan performans puanları** arasındaki korelasyon katsayısı nokta **0.902** eşittir

Çoklu Regresyon

Çoklu Korelasyon

- Çoklu korelasyon katsayısının kestirimi hem örneklem büyüklüğüne (n) hem de bağımsız değişkenlerin sayısına (k) bağlıdır.

- Bağımlı değişken ile bağımsız değişkenler arasında hiç ilişki yoksa, R değerinin sıfıra yakın olması beklenir ancak R'nin beklenen değeri rastgele bir veri için $k/(n-1)$'dir.

Çoklu Regresyon

- Örneğin, örneklem büyüklüğünün **50**, bağımsız değişken sayısının **2** olduğu bir durumda, bağımlı değişken ile bağımsız değişkenler arasında hiç ilişki yoksa, R değeri **0.04** olacaktır, **0** değil.

- Bu nedenle büyük örnekleme sahip olmak önemlidir. Her bağımsız değişken için **en az 10** gözlem önerilmektedir. Bir başka öneri de örneklem büyüklüğünün **bağımsız değişken sayısından en az 50 fazla olması** yönündedir.

- Bir çalışmada, **tek bir bağımsız değişken bulunduğunda**, **0.80** güce sahip olmak için **0.30** evren korelasyonunun **124** birey gerektireceği belirtilmiştir. **Beş bağımsız** değişken bulunduğundaysa, örneklem büyüklüğünün **187** olması gerekmektedir.

Çoklu Regresyon

Çoklu Korelasyon

```
model <- lm(Performans ~ Motivasyon + Kaygi,data=performans)
glance(model)[,1]
```

```
## # A tibble: 1 × 1
##   r.squared
##       <dbl>
## 1     0.814
```

R^2 değeri **belirlilik katsayısı** (coefficient of determination) olup bağımlı değişkenin gözlenen ve yordanan değerleri arasındaki korelasyonun karesi alınarak hesaplanır. Bu değer bağımlı değişkendeki varyansın model tarafından açıklanan oranını ifade eder. Diğer bir ifadeyle **bağımlı değişkenin varyansının bağımsız değişkenlerin en iyi doğrusal kombinasyonu** ile paylaşılan oranını ifade eder. **Performans puanlarındaki varyansın yaklaşık %81'i öğrencilerin motivasyon ve kaygı puanları** tarafından açıklanabilir.

Çoklu Regresyon

- Modele yeni bir bağımsız değişken eklendiğinde, R^2 degeri artar, sadece **şans eseri olsa bile**. Böylece daha fazla bağımsız değişken içeren model sadece daha fazla bağımsız değişken içerdiçi için veriye daha iyi uyum sağlıyor gibi gözükebilir.

- Bu etkiyi gidermek icin $\mathbf{adj\ R\hat{}2}$ değeri hesaplanabilir.

- $\mathbf{adj\ R\hat{}2}$değeri, R^2 değerinin modeldeki bağımsız değisken sayısı için modifiye edilmiş versiyonudur. $\mathbf{adj\ R\hat{}2}$ değeri yeni eklenen bağımsız değişken modeli şans eseri beklenenden daha fazla gelştirirse artar, daha az geliştirirse azalır.

- $\mathbf{adj\ R\hat{}2}$ değeri, eksi değerler alabilir ancak genellikle artı değerler alır. Her zaman R^2 değerinden daha düşüktür.

Çoklu Regresyon

R^2 değeri, n gözlemlerin sayısı, k bağımsız değişkenlerin sayısı olmak üzere, aşağıdaki eşitlikle hesaplanabilir.

$$R^2_{adj} = R^2 - \frac{k - (1 - R^2)}{n - k - 1}$$

$$R^2_{adj} = 0.814 - \frac{2 - (1 - 0.814)}{15 - 2 - 1} = 0.783$$

```
glance(model)[,2]
```

```
## # A tibble: 1 × 1
##   adj.r.squared
##           <dbl>
## 1         0.783
```

$adjR^2$ evrende gerçek korelasyonun karesinin daha az yanlı kestirimi olsa da, çoğunlukla R^2 değeri rapor edilir.

Çoklu Regresyon

- **Kestirimin standart hatası** (standard error of the estimation), modeldeki artıkların karelerinin toplamının,

- $n-p$ (n örneklem büyüklüğü ve p modeldeki parametrelerin sayısı) ile bölünmesiyle elde edilen bölümün kareköküdür.

```
res <- model$residuals
sqrt(sum((res - mean(res))^2/(length(res)-3)))
```

```
## [1] 3.65
```

```
glance(model)[,3]
```

```
## # A tibble: 1 × 1
##   sigma
##   <dbl>
## 1  3.65
```

Çoklu Regresyon

- Modelin veriye iyi uyup uymadığının test edilmesinde kullanılacak **F** değeri varyans analizi sonucunda elde edilir.

- Regresyonun anlamlılığının test edildiği varyans analizinde, birlikte ele alınan bir grup bağımsız değişkenin (motivasyon ve kaygı gibi) en iyi doğrusal kombinasyonu ile bağımlı değişken (performans gibi) arasında **korelasyon yoktur sıfır hipotezi** test edilir.

- İstatistiksel olarak **anlamlı etki**, **evrende çoklu korelasyon katsayısının sıfırdan farklı olduğu anlamına** gelir.

Çoklu Regresyon

```r
glance(model)[,4:6]
```

```
## # A tibble: 1 × 3
##   statistic   p.value    df
##       <dbl>     <dbl> <dbl>
## 1      26.2 0.0000420     2
```

- F istatistiği 26.2 değerine eşittir ve istatistiğe ilişkin $p < 0.001$. Bu olasılık 0.05'ten küçük olduğundan, sıfır hipotezi reddedilir.

- Bu sonuç **motivasyon ve kaygı değişkenlerinin ikisi birlikte kullanıldığında, çoklu korelasyon katsayısının anlamlı olarak sıfırdan büyük olduğunu** ifade etmektedir. Diğer bir ifadeyle, motivasyon ve kaygı değişkenleri performansı istatistiksel olarak anlamlı bir şekilde yordamaktadır.

- **Regresyon modeli veriye iyi uyum sağlamaktadır.**

Çoklu Regresyon

```r
library(knitr)
tidy(model) %>% kable()
```

term	estimate	std.error	statistic	p.value
(Intercept)	1.744	5.096	0.342	0.738
Motivasyon	0.686	0.098	6.975	0.000
Kaygi	-0.607	0.207	-2.936	0.012

Performans puanlarındaki farklılıkların **bir kısmı motivasyon puanlarındaki farklılıklardan**, **bir kısmı ise kaygı puanlarındaki farklılıklardan** kaynaklanmaktadır

Çoklu Regresyon

- **Performansın** sadece **kaygıdan yordandığı basit regresyon analizi** gerçekleştirilirse, yordanan puanlar ve gözlenen puanlar arasındaki fark **(artıkPER1)**, **performansın kaygıdan yordanamayan kısmı** olacaktır.

```
artıkPER1 <- lm(Performans ~  Kaygi,data=performans)$residuals
```

- **Motivasyonun** sadece **kaygıdan yordandığı basit regresyon analizi** gerçekleştirilirse, yordanan puanlar ve gözlenen puanlar arasındaki fark **(artıkMOT)**, **motivasyonun kaygıdan yordanamayan kısmı** olacaktır.

```
artıkMOT <- lm(Motivasyon  ~  Kaygi,data=performans)$residuals
```

- Böylece **artıkPER1** ve **artıkMOT** olarak adlandırılan artık puanlar kaygıdan bağımsız olacaktır.

Çoklu Regresyon

- Diğer bir ifadeyle, **kaygı ilişkilerinde** herhangi bir rol oynamayacaktır. **artıkPER1** puanları **artıkMOT** puanlarından yordanırsa, **artıkMOT** puanlarına ilişkin eğim katsayısı **0.686** olarak kestirilecektir. Bu değer, **öğrencilerin kaygı düzeyleri kontrol altına** alındıktan sonra, **motivasyon düzeylerindeki bir birimlik** artışın matematikteki **performans düzeylerini 0.686** birim artırmaya eğilimli olduğunu önermektedir

```
round(lm(artıkPER1  ~  artıkMOT,data=data.frame(artıkPER1,artıkMOT))$coefficients,3)
```

```
## (Intercept)     artıkMOT
##       0.000        0.686
```

Çoklu Regresyon

- $$B_{Y_{12}} = \frac{r_{Y1} - r_{Y2}r_{12}}{1 - r_{12}^2}\frac{sd_Y}{sd_1}$$

- $$B_{Y_{12}} = \frac{(0.824) - (-0.241)(0.147)}{1 - (0.022)}\frac{7.827}{10.025} = 0.879 * 0.780 = 0.686$$

- Bu değer, **kaygı puanı kontrol altına alındıktan sonra**, **motivasyon** puanlarındaki **bir birimlik artışın** öğrencilerin **matematik performansından 0.686** birim artmaya eğilimi olduğunu önermektedir.

```r
attach(performans)
((cor(Motivasyon,Performans) - cor(Kaygi ,Performans)*cor(Motivasyon,Kaygi))/
   (1-cor(Motivasyon,Kaygi)^2))*(sd(Performans)/sd(Motivasyon))
```

```
## [1] 0.686
```

Çoklu Regresyon

- **Performansın** sadece **motivasyondan** yordandığı basit regresyon analizi gerçekleştirilirse, yordanan puanlar ve gözlenen puanlar arasındaki fark **(artıkPER2)**, **performansın motivasyondan** yordanamayan kısmı olacaktır.

```
artıkPER2 <- lm(Performans ~  Motivasyon ,data=performans)$residuals
```

- **Kaygının** sadece **motivasyondan** yordandığı basit regresyon analizi gerçekleştirilirse, yordanan puanlar ve gözlenen puanlar arasındaki fark **(artıkKAY)**, **kaygının motivasyondan** yordanamayan kısmı olacaktır.

```
artıkKAY <- lm(Kaygi ~  Motivasyon ,data=performans)$residuals
```

Çoklu Regresyon

- Böylece **artıkPER2** ve **artıkKAY** olarak adlandırılan artık puanlar motivasyondan bağımsız olacaktır. Diğer bir ifadeyle, **motivasyon ilişkilerinde herhangi bir rol** oynamayacaktır.

- **artıkPER2** puanları **artıkKAY** puanlarından yordanırsa, **artıkKAY** puanlarına ilişkin eğim katsayısı **-0.607** olarak kestirilecektir. Bu değer, **öğrencilerin motivasyon düzeyleri kontrol altına alındıktan sonra**, kaygı düzeylerindeki **bir birimlik** artışın matematikteki performans düzeylerini **0.607** birim azaltmaya eğilimli olduğunu önermektedir

```r
round(lm(artıkPER2  ~  artıkKAY,data=data.frame(artıkPER2,artıkKAY))$coefficients,3)
```

```
## (Intercept)    artıkKAY
##       0.000      -0.607
```

Çoklu Regresyon

$$B_{Y_{21}} = \frac{r_{Y2} - r_{Y1}r_{12}}{1 - r_{12}^2}\frac{sd_Y}{sd_2}$$

$$B_{Y_{12}} = \frac{(-0.241) - (0.824)(0.147)}{1 - (0.022)}\frac{7.827}{4.769} = (-0.370) * (1.641) = -0.607$$

Bu değer, **motivasyon puanı kontrol altına alındıktan sonra**, kaygı puanlarındaki **bir birimlik** artışın öğrencilerin matematik performansından **0.607** birim azaltmaya eğilimli olduğunu önermektedir.

```
((cor(Kaygi,Performans)- cor(Motivasyon ,Performans)*cor(Motivasyon,Kaygi))/
  (1-cor(Motivasyon,Kaygi)^2))*(sd(Performans)/sd(Kaygi))
```

```
## [1] -0.607
```

Çoklu Regresyon

$$B_0 = M_Y - B_{Y12} * M_1 - B_{Y21} * M_2$$

$$B_0 = 18.176 - (0.686) * (39.933) - (-0.607) * (18.701) = 1.744$$

```
mean(Performans)-
  (model$coefficients[2]*mean(Motivasyon))-
  (model$coefficients[3]*mean(Kaygi))
```

```
## Motivasyon
##       1.74
```

Bu değer hem motivasyon puanı hem de kaygı puanı 0'a eşit olduğunda **yordanan performans puanıdır.**

Çoklu Regresyon

Böylece yordanan performans puanı

$$Y_{performans_i} = 1.744 + 0.686 * X_{motivasyon_i} - 0.607 * X_{kaygi_i}$$

Çoklu Regresyon

```
library(QuantPsyc)
lm.beta(model)
```

```
## Motivasyon      Kaygi
##      0.879     -0.370
```

- Çoklu regresyon eşitliğini elde etmeden önce değişkenlerin her biri **standartlaştırılırsa** (değişkenlerin her birinin ortalaması 0, standart sapması 1 olacak şekilde ayarlanırsa), sonuçlar **standart sapma** birimlerince ifade edilir.

- Böylece örnekte standartlaştırılmış değişkenler kullanıldığında, yordanan standartlaştırılmış performans düzeyleri aşağıdaki eşitlikle hesaplanabilir:

$$Y_{Zperformans_i} = 0.879X_{Zmotivasyon_i} + -0.370X_{Zkaygi_i}$$

- Değişkenler standartlaştırıldığında, **kesişim katsayısı 0** olacaktır ve eşitlikte **gösterilmeyecektir.**

Çoklu Regresyon

- Motivasyon için **standartlaştırılmış eğim katsayısı** $\beta_{motivasyon}$ **0.879** değerine eşittir.
- Bu değer, **kaygı puanı kontrol altına alındıktan** sonra, **motivasyon** puanındaki **bir standart sapmalık artışın** öğrencilerin matematikteki performans puanlarını **0.879 standart sapma** artırmaya eğilimli olduğunu önermektedir.
- Benzer şekilde, **kaygı için standartlaştırılmamış eğim katsayısı** β_{kaygi} -0.370 değerine eşittir. Bu değer, **motivasyon puanı kontrol altına** alındıktan sonra, kaygı puanındaki **bir standart sapmalık artışın** öğrencilerin matematikteki performans puanlarını **0.370 standart sapma** azaltmaya eğilimli olduğunu önermektedir.

Çoklu Regresyon

- Motivasyonun standartlaştırılmış eğim katsayısının mutlak değeri, kaygının standartlaştırılmış eğim katsayının mutlak değerinden daha büyük olduğundan, motivasyonun öğrencilerin matematikteki performanslarını yordamada kaygıya göre daha önemli bir yordayıcı olduğu söylenebilir.

$$|0.879| > |-0.307|$$

Çoklu Regresyon

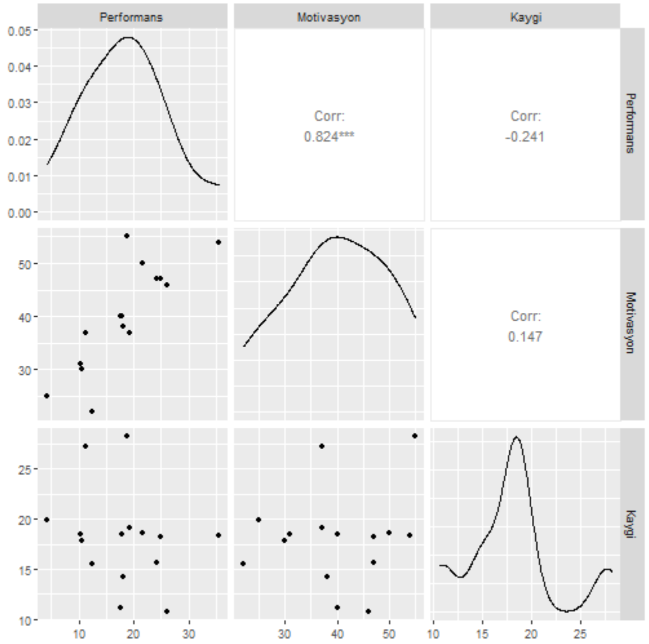
- Standartlaştırılmış eğim katsayılarının bağıl büyüklükleri "önemin" en iyi göstergeleri olmasa da, yorumlanmaları kolaydır ve regresyon analizlerinin yürütülmesinde yararlanılan bilgisayar programlarının çoğu tarafından yazdırılır.

- Ancak bağımsız değişkenlerin **standartlaştırılmamış eğim katsayılarını karşılaştırmak uygun değildir.**

- NOT: Bağımsız değişkenler arasında korelasyon olduğunda, **standartlaştırılmış ğim katsayısı** bağımlı değişken ile bağımsız değişken arasındaki **korelasyon katsayısı değildir.**

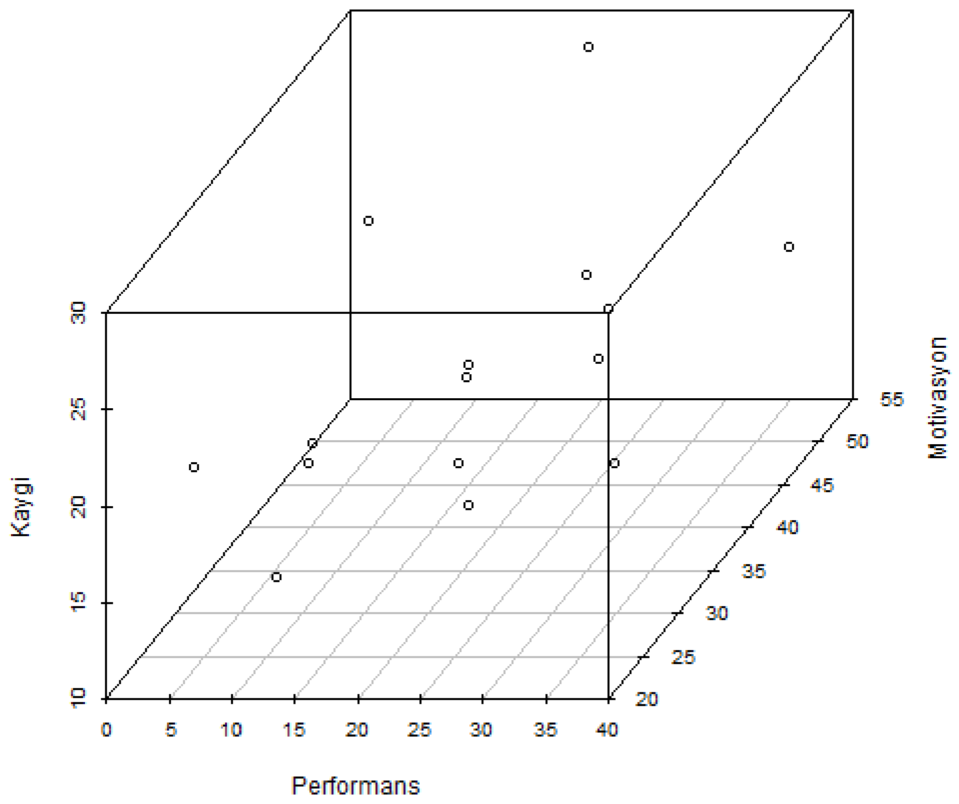
Çoklu Regresyon

```r
library(GGally)
ggpairs(performans[,1:3])
```



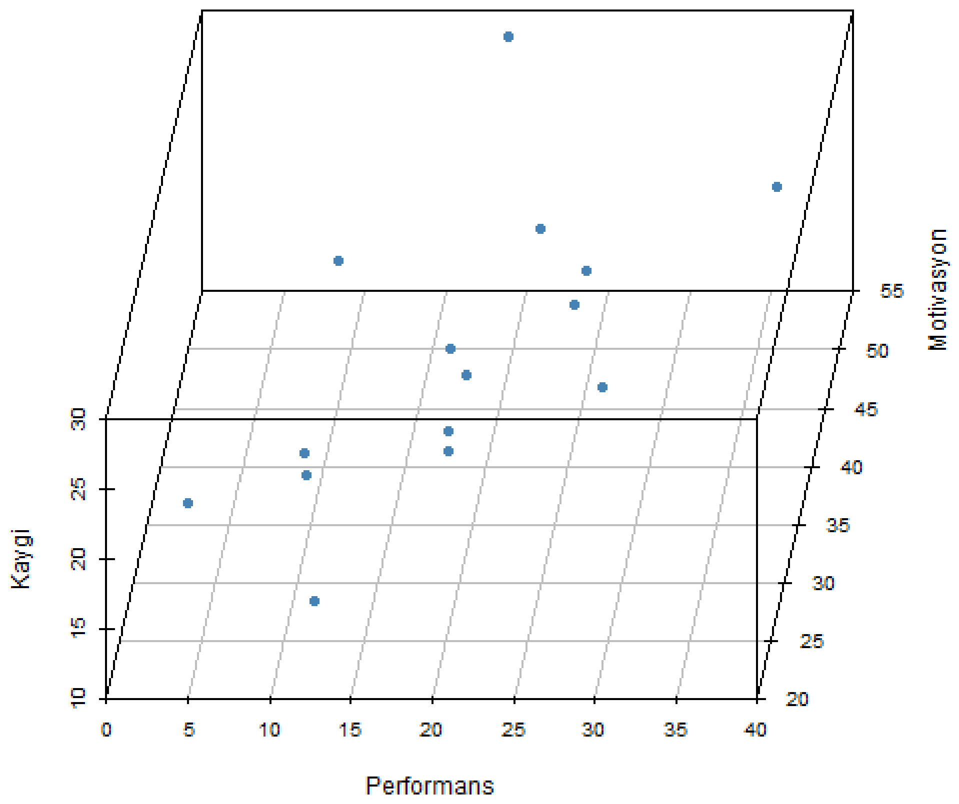
Çoklu Regresyon

```r
library(scatterplot3d)
scatterplot3d(performans[,1:3])
```



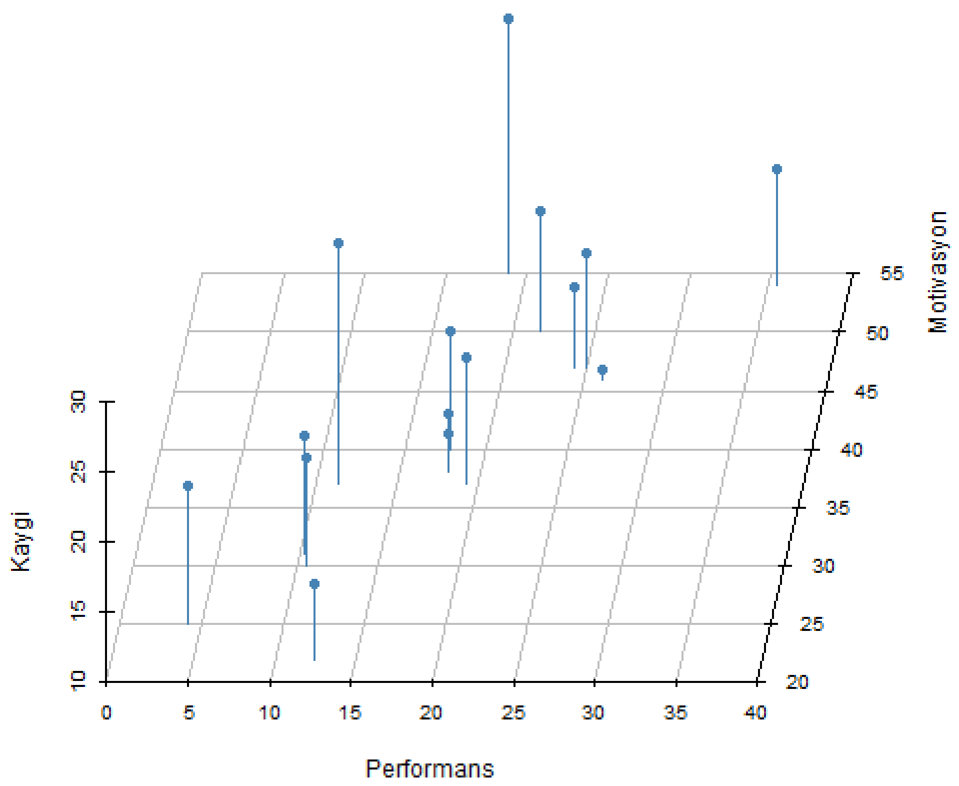
Çoklu Regresyon

```
library(scatterplot3d)
scatterplot3d(performans[,1:3],pch = 16, color="steelblue", angle=75)
```



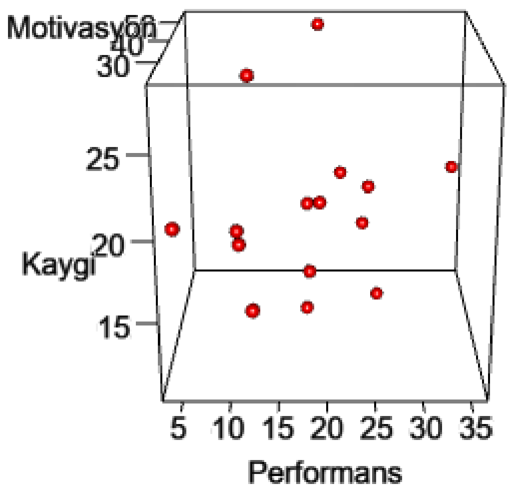
Çoklu Regresyon

```
scatterplot3d(performans[,1:3],pch = 16, color="steelblue",angle=75,box = FALSE,type = "h")
```



Çoklu Regresyon

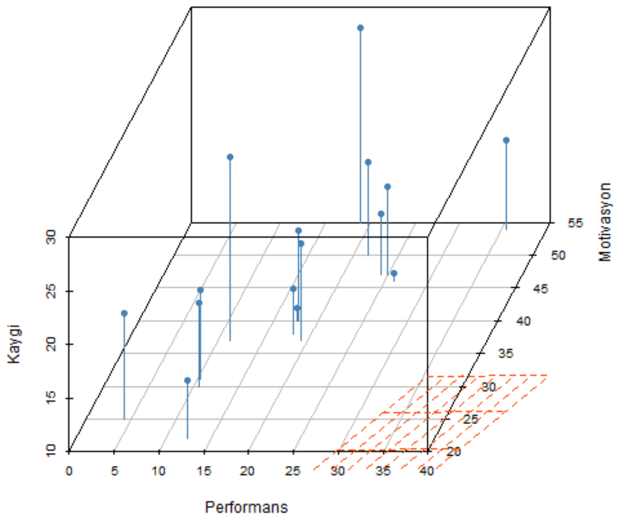
```
library(rgl)
plot3d(Performans, Motivasyon, Kaygi,
xlab = "Performans", ylab = "Motivasyon", zlab = "Kaygi", type = "s",size = 1.5,col = "red")
rglwidget()
```



Çoklu Regresyon

```r
p <- scatterplot3d(performans[,1:3], angle=55,type='h',
                   pch = 16, color = "steelblue")

# add a plane representing the fit of the model
p$plane3d(model, col='orangered')
```



Çoklu Regresyon

```
# augment(model,data=performans)[,5]
model$fitted.values
```

```
##     1     2     3     4     5     6     7     8     9    10    11    12
##  7.46  6.87 24.75 22.95 26.80 22.42 11.84 22.38 15.56 18.01 27.69 19.20
##    13    14    15
## 10.66 24.50 11.55
## attr(,"format.spss")
## [1] "F8.2"
## attr(,"display_width")
## [1] 9
```

```
model$residuals
```

```
##      1      2      3      4      5      6      7      8      9     10
##  4.910 -2.757 -3.123  1.972 -0.716 -4.715 -1.565 -3.485  3.642 -0.262
##     11     12     13     14     15
##  7.809 -1.160  0.701 -0.246 -1.004
## attr(,"format.spss")
## [1] "F8.2"
## attr(,"display_width")
## [1] 9
```

- Öğrencilerin **standratlastırılmamış yordanan matematik performans düzeyleri** ve **standartlaştırılmamış artıkları** modelden çekilebilir.
- Örneğin, ilk öğrenci için standratlastırılmamış yordanan değer yaklaşık **7.46**, artık ise yaklaşık **4.910**'tir.

Çoklu Regresyon

```r
library(outliers)
model$fitted.values %>% scores(type = "z")
```

```
##       1       2       3       4       5       6       7       8       9
## -1.5176 -1.6014  0.9311  0.6760  1.2219  0.6005 -0.8978  0.5959 -0.3708
##      10      11      12      13      14      15
## -0.0231  1.3478  0.1452 -1.0648  0.8960 -0.9390
## attr(,"format.spss")
## [1] "F8.2"
## attr(,"display_width")
## [1] 9
```

```r
model$residuals %>% scores(type = "z")
```

```
##       1       2       3       4       5       6       7       8       9
##  1.4528 -0.8158 -0.9242  0.5836 -0.2118 -1.3950 -0.4631 -1.0313  1.0775
##      10      11      12      13      14      15
## -0.0775  2.3106 -0.3432  0.2073 -0.0729 -0.2970
## attr(,"format.spss")
## [1] "F8.2"
## attr(,"display_width")
## [1] 9
```

- Öğrencilerin **standratlastırılmış yordanan matematik performans düzeyleri** ve **standartlaştırılmış artıkları** modelden çekilebilir.

Çoklu Regresyon

- Regresyon katsayılarından her birinin istatistiksel olarak sıfırdan farklı olup olmadığı test edilebilir. Bu durumda regresyon katsayılarına ilişkin test edilecek sıfır hipotezleri aşağıdaki gibidir:

$$H_0 : \beta_1 = 0$$

- Kaygı düzeyleri eşit olan öğrenciler için motivasyon düzeylerindeki farklılıklar performans düzeylerinde farklılığa yol açar mı?

$$H_0 : \beta_2 = 0$$

- Motivasyon düzeyleri eşit olan öğrenciler için kaygı düzeylerindeki farklılıklar performans düzeylerinde farklılığa yol açar mı?

Çoklu Regresyon

Hipotez testlerine ilişkin **t istatistiği**, standartlaştırılmamış regresyon katsayılarının standart hatalarına bölünmesi ile hesaplanır.

```
tidy(model) %>% kable()
```

term	estimate	std.error	statistic	p.value
(Intercept)	1.744	5.096	0.342	0.738
Motivasyon	0.686	0.098	6.975	0.000
Kaygi	-0.607	0.207	-2.936	0.012

```
0.686 /0.0984
```

```
## [1] 6.97
```

- Motivasyona ilişkin eğim için testin, olasılık değeri ($p < 0.001$) 0.05'ten daha küçük olduğundan, anlamlı olarak sıfırdan farklı olduğu görülmektedir.

Çoklu Regresyon

- Hipotez testlerine ilişkin t istatistiği standartlaştırılmamış regresyon katsayılarının standart hatalarına bölünmesi ile hesaplanır.
- Kaygıya ilişkin eğim de anlamlıdır **($t = -2.936$, $p = 0.012$)**, öğrencilerin motivasyon düzeylerindeki farklılıklar kontrol altına alınsa bile, öğrencilerin kaygı düzeyleri performans düzeylerinde fark yapmaktadır ve kaygı düzeyi negatif bir etkiye sahiptir.

Çoklu Regresyon

- Hipotez testlerine ilişkin t istatistiği **standartlaştırılmamış regresyon katsayılarının standart hatalarına** bölünmesi ile hesaplanır.
- Regresyon katsayısının **standart hatası tekrarlanan örneklemlerde istatistiğin değişkenliğini** belirtir.
- İki regresyon katsayısı ile ilgili **p** değerleri 0.05 alfa düzeyinden daha küçüktür, bu nedenle her iki bağımsız değişken de öğrencilerin matematikteki performanslarını yordamada **istatistiksel olarak anlamlıdır.**

Çoklu Regresyon

- Çoklu regresyon modelini bir **yol şeması** ile sunmak oldukça kullanışlıdır.

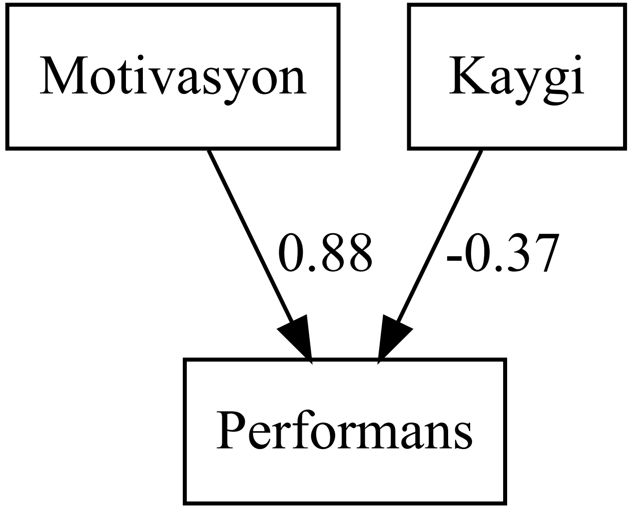
```r
# path model
library(lavaan)
library(lavaanPlot)
model_1 <- 'Performans ~  Motivasyon + Kaygi'
fit1 <- sem(model_1, data = performans)
coef(fit1)
```

```
##   Performans~Motivasyon       Performans~Kaygi Performans~~Performans
##                   0.686                 -0.607                 10.661
```

Çoklu Regresyon

Standart çözüm

```
lavaanPlot(model = fit1, coefs = TRUE, stand = TRUE, sig = 0.05)
```

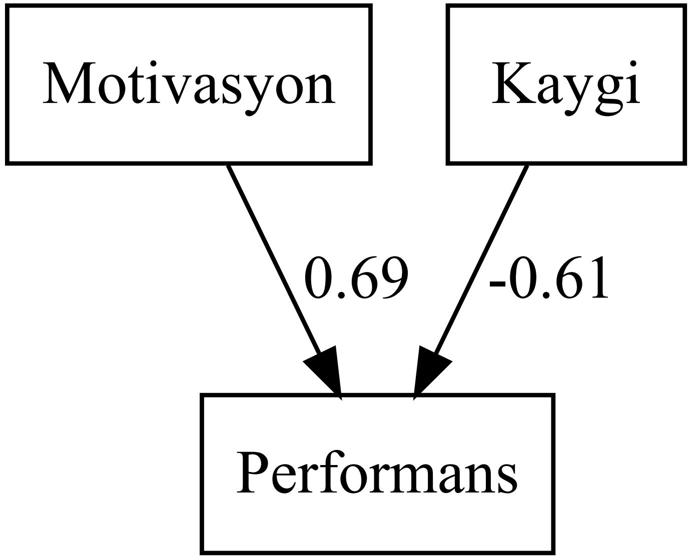


Çoklu Regresyon

Standart olmayan çözüm:

Not: Araştırmacılar özellikle ortalamaların yapısı ile ilgilenmedikleri sürece yol şemasında b_0 gösterilmez.

```r
lavaanPlot(model = fit1,
coefs = TRUE, stand = FALSE, sig = 0.05)
```



Çoklu Regresyon

- Öğrencilerin matematik performanslarının düzeyini motivasyon ve kaygı düzeylerinden yordamak için çoklu regresyon analizi gerçekleştirilmiştir. Genel regresyon istatistiksel olarak anlamlıdır $\boldsymbol{F_{[2,12]} = 26.188, p < .001}$

- $\boldsymbol{R^2 = 0.814}$ Öğrencilerin hem motivasyonlarının düzeyi ($\boldsymbol{b_1 = 0.686}$) hem kaygılarının düzeyi($\boldsymbol{b_2 = -0.607}$) matematik performanslarının düzeyinin istatistiksel olarak anlamlı yordayıcılarıdır,
$\boldsymbol{t = 6.975; p < .001, t = -2.936; p = .012}$

Regresyon Eşitliğinin Oluşturulması

Aşamalı (Stepwise) Regresyon

- Bir regresyon modeline dahil edilebilecek çok sayıda değişken bulunduğunda, bu değişkenlerden **en uygun** regresyon eşitliğinin oluşturulması için değişken seçiminde çeşitli yöntemler vardır. Bu yöntemlerden birisi aşamalı **stepwise** regresyondur.

- Aşamalı regresyon yöntemi **her bağımsız değişkenin regresyon modeline** katkısının incelenmesini sağlar.

- Bu yönteme göre önce **bağımlı değişkenle en yüksek korelasyona** sahip bağımsız değişken seçilerek basit regresyon modeli kurulur.

Regresyon Eşitliğinin Oluşturulması

Aşamalı (Stepwise) Regresyon

- Birinci regresyon eşitliğinden kalan **hata varyansının istatistiksel olarak anlamlı kısmını en çok açıklayan bağımsız değişkeni** bulmak için **kısmi korelasyon** katsayıları incelenir ve en yüksek **kısmi korelasyon katsayısına** sahip bağımsız değişken modele eklenir.

- İki bağımsız değişken ile regresyon eşitliği yeniden hesaplanır ve **eklenen değişkenin modele anlamlı katkısı olup olmadığı** test edilir. Bu işlem modele anlamlı katkı sağlayacak değişken kalmayana kadar devam eder.

Aşamalı regresyon

```
sadece_kesisim  <- lm(Performans ~ Motivasyon, data=performans)
glance(sadece_kesisim)
```

```
## # A tibble: 1 × 12
##   r.squared adj.r.s…¹ sigma stati…² p.value    df logLik   AIC   BIC devia…³
##       <dbl>     <dbl> <dbl>   <dbl>   <dbl> <dbl>  <dbl> <dbl> <dbl>   <dbl>
## 1     0.680     0.655  4.60    27.6 1.56e-4     1  -43.1  92.2  94.3    275.
## # … with 2 more variables: df.residual <int>, nobs <int>, and abbreviated
## #   variable names ¹adj.r.squared, ²statistic, ³deviance
```

```
tum  <- lm(Performans ~  Motivasyon + Kaygi, data=performans)
glance(tum)
```

```
## # A tibble: 1 × 12
##   r.squared adj.r.s…¹ sigma stati…² p.value    df logLik   AIC   BIC devia…³
##       <dbl>     <dbl> <dbl>   <dbl>   <dbl> <dbl>  <dbl> <dbl> <dbl>   <dbl>
## 1     0.814     0.783  3.65    26.2 4.20e-5     2  -39.0  86.1  88.9    160.
## # … with 2 more variables: df.residual <int>, nobs <int>, and abbreviated
## #   variable names ¹adj.r.squared, ²statistic, ³deviance
```

Aşamalı regresyon

- Birinci modelde **motivasyon tek yordayıcıdır** ve performans ile korelasyonu **0.824'tür** $R = 0.824$. Motivasyon tek başına performans puanlarındaki varyansın yaklaşık **%68'ini** $R^2 = 0.680$ açıklamaktadır.

- Modele **kaygının yordayıcı olarak eklenmesiyle** korelasyon **0.902**'ye $R = 0.902$ yükselmiştir. Motivasyon ve kaygı birlikte performans puanlarındaki varyansın yaklaşık **%81**'ini $R^2 = 0.814$ açıklamaktadır.

Aşamalı regresyon

- Modele kaygının eklenmesiyle R^2 değişimi (R Square Change) 0.134'tür. R^2 değerindeki bu değişim kaygının eklenmesiyle açıklanan varyans oranında **%13**'lük bir artış olduğu anlamındadır. R^2 değişimi F testi (F Change) ile test edilmiştir ve F değerindeki değişim istatistiksel olarak anlamlıdır $p = 0.012$ Dolayısıyla modele eklenen kaygı değişkeni yordamayı anlamlı olarak geliştirmiştir.

```
asamali <- step(sadece_kesisim, direction='forward', scope=formula(tum), trace=0)
tidy(anova(tum,sadece_kesisim))
```

```
## # A tibble: 2 × 7
##   term                              df.res…¹   rss    df sumsq stati…² p.value
##   <chr>                                <dbl> <dbl> <dbl> <dbl>   <dbl>   <dbl>
## 1 Performans ~ Motivasyon + Kaygi         12  160.    NA    NA     NA     NA
## 2 Performans ~ Motivasyon                 13  275.    -1 -115.    8.62  0.0125
## # … with abbreviated variable names ¹df.residual, ²statistic
```

Aşamalı regresyon

- Modelde tek bir yordayıcı (motivasyon) varken, korelasyon **0.824**'tür ve sıfır hipotezi doğruysa bu kadar yüksek bir korelasyon elde etme olasılığı $p < 0.001$. Bu olasılık 0.05'ten küçük olduğundan, korelasyonun anlamlı olarak sıfırdan büyük olduğu söylenebilir.

- Modele motivasyon değişkeninin yanı sıra kaygi değişkeni de yordayıcı olarak eklendiğinde, çoklu korelasyon **0.902**'dir ve sıfır hipotezi doğruysa bu kadar yüksek bir korelasyon elde etme olasılığı $p < 0.001$. Bu olasılık 0.05'ten küçük olduğundan, çoklu korelasyonun anlamlı olarak sıfırdan büyük olduğu söylenebilir.

Aşamalı regresyon

- Birinci modelde motivasyon tek yordayıcıdır. Bu modelde motivasyona ilişkin standartlaştırılmamış eğim katsayısı **0.644** olarak kestirilmiş olup kestirimin standart hatası **0.123**'tür. Bu katsayı $p < .05$'te anlamlıdır. Standartlaştırılmış eğim katsayısı 0.824 olarak kestirilmiştir ve bu değer motivasyon ile performansarasındaki korelasyondur.

- İkinci modelde **motivasyon ve kaygı** yordayıcılardır. Bu modelde motivasyona ilişkin standartlaştırılmamış eğim katsayısı **0.686** olarak kestirilmiş olup kestirimin standart hatası **0.098**'dir.

Aşamalı regresyon

- Öğrencilerin kaygı düzeyi kontrol altına alındığında, artan motivasyon düzeyi daha yüksek performans puanları ile ilişkilidir. Bu katsayı $p < .05$'te anlamlıdır. Standartlaştırılmış eğim katsayısı 0.879 olarak kestirilmiştir. Kaygıya ilişkin standartlaştırılmamış eğim katsayısı -0.607 olarak kestirilmiş olup kestirimin standart hatası 0.207'dir. Öğrencilerin motivasyon düzeyi kontrol altına alındığında, artan kaygı düzeyi daha düşük performans puanları ile ilişkilidir.

- Kaygının modele eklenmesi korelasyonu çok fazla artırmasa da istatistiksel olarak anlamlı bir yordayıcıdır. Standartlaştırılmış eğim katsayısı -0.370 olarak kestirilmiştir.

Etkili Gözlemlerin Belirlenmesi

- Etkili gözlemler (influential observations) regresyon sonuçları üzerinde **orantısız etkisi** olan bütün gözlemleri içerir.

- Bu **aşırı değerler regresyon doğrusunu kendilerine doğru çekerek** modelin katsayıları üzerinde anlamlı etkileri olan değerlerdir.

- Bu gözlemlerin belirlenmesinde kullanılan bazı istatistikler şunlardır:

- **Distance (Uzaklık)**
- **Leverage (hi)**
- **Influence (Etki)**
- **Cook's D**

Etkili Gözlemlerin Belirlenmesi

Distance (Uzaklık):

- **Bağımlı değişkenlerdeki potansiyel uç değerlerin** belirlenmesinde kullanışlı bir istatistiktir.
- **Uzaklık için en yaygın ölçüm artıktır**.
- Artık herhangi **bir nokta ve regresyon eğrisi arasındaki dikey uzaklığı** ölçer. Bu noktalar **rastgele hatayı temsil edebilir**, Veri yanlış kodlanmış olabilir veya veri setine ait olmayan olağan dışı durumları yansıtabilir.

Etkili Gözlemlerin Belirlenmesi

Leverage (hi):

- **Bağımsız değişkenlerdeki potansiyel uç değerlerin belirlenmesinde** kullanışlı bir istatistiktir.
- Levarage bir gözlemin bir bağımsız değişkene, X_j, göre **olağan dışı olma derecesini** ölçer.
- Leverage için olası değerler, N gözlemlerin sayısı olmak üzere, **1/N** ile **1.0** arasında değişir.
- Ortalama leverage puanı, **p** bağımsız değişken sayısı ve **N** gözlem sayısı olmak üzere, **(p +1)/N** eşitliği ile hesaplanabilir.
- **Yüksek leverage değerine** sahip gözlemler ortalama değerden 2 veya 3 kat daha yüksek leverage puanlarına sahip olacaktır.

Etkili Gözlemlerin Belirlenmesi

Influence (Etki):

- Etkili bir gözlem **uzaklık ve/veya leverage için yüksek değere** sahip olan ve **modelin kesişim ve eğim katsayılarını anlamlı** olarak etkileyen bir gözlemdir.
- Bu gözlemin varlığı veya yokluğu regresyon yüzeyinin yerini önemli ölçüde değiştirecektir.
- Uzaklık ve/veya leverage için yüksek değere sahip gözlemlerin regresyon üzerinde önemli bir etkisi olmayabilir. Bir gözlemin etkide yüksek olması için **hem uzaklık hem de leverage için yüksek değerlere** sahip olması gerekir.

Etkili Gözlemlerin Belirlenmesi

Cook's D

- Etkinin en yaygın ölçümü **Cook's D** olarak bilinir.

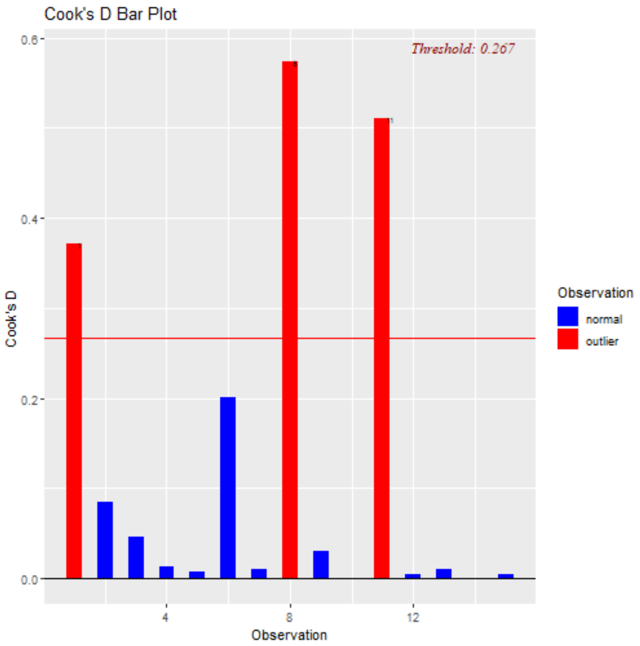
- Cook's D **i** gözlemi veriden çıkarılıp analiz yeniden gerçekleştirilirse, b_j katsayısındaki değişikliğin karesinin toplamının bir fonksiyonudur.

- Her gözlem için hesaplanabilir. Her gözlem için bu değer, **N** gözlemlerin sayısı olmak üzere **4/N** ile karşılaştırılabilir. **4/N** üzerindeki değerler problem olabilecek gözlemlere işaret eder.

- **Cook's D** etkinin genel bir ölçümü olarak düşünülebilir. Gözlemin eklenmesiyle her katsayının nasıl değiştiğini ölçen daha spesifik bir ölçüm ele alınabilir. Bu ölçüm **DFBETA** olarak adlandırılır ve her gözlem için hesaplanabilir. (kritik değer $2/\sqrt{n}$

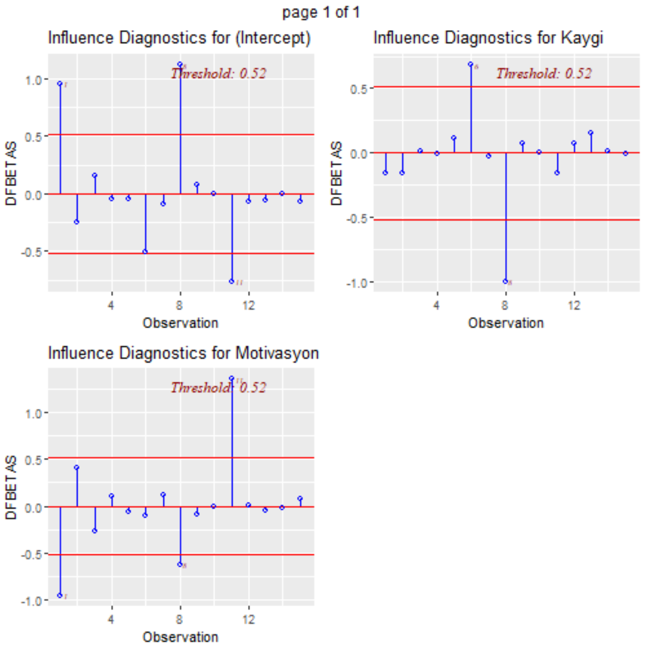
Cook's D

```
library(olsrr)
ols_plot_cooksd_bar(model)
```

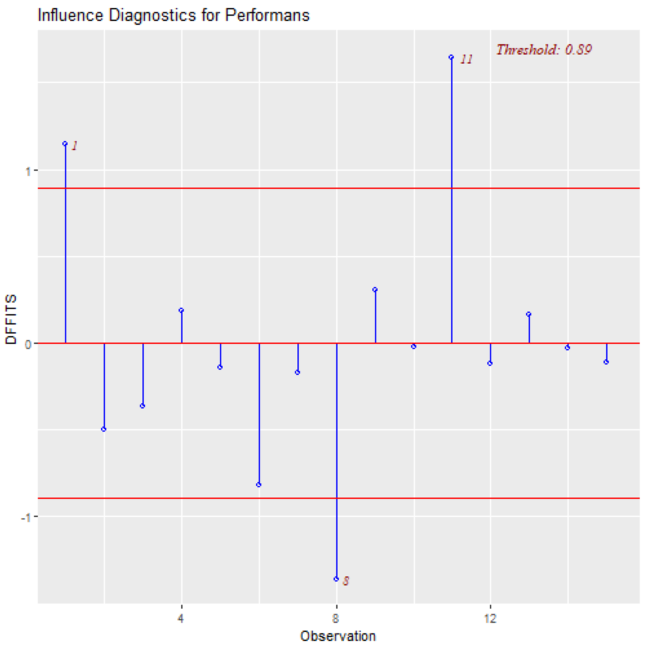


DFBETA

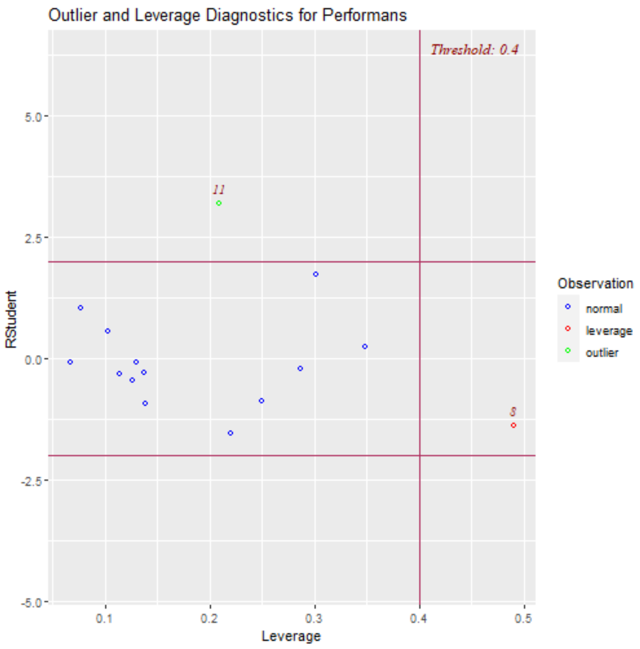
```
ols_plot_dfbetas(model)
```



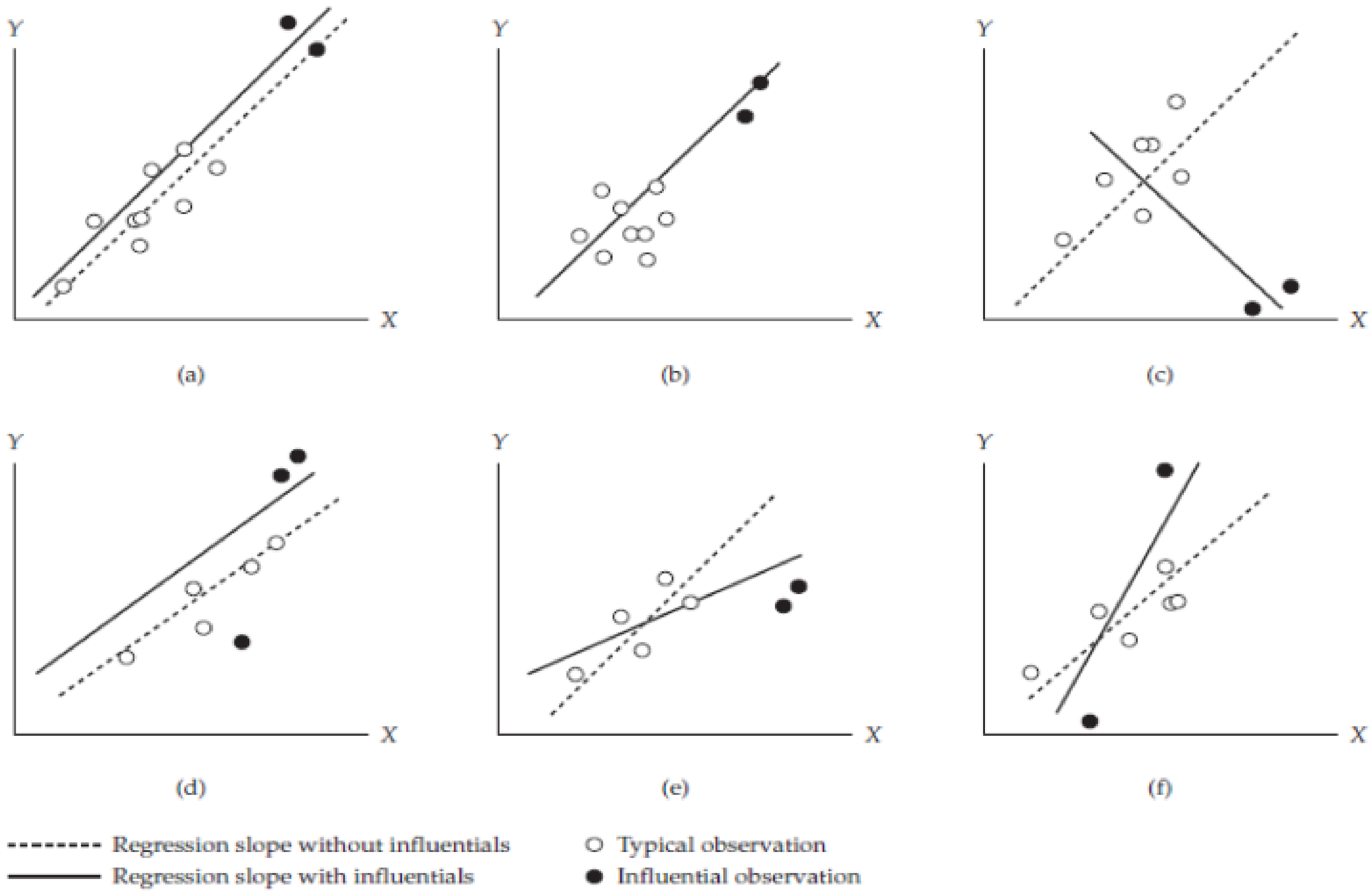
```
ols_plot_dffits(model)
```



```
ols_plot_resid_lev(model)
```



Etkili Gözlemlerin Belirlenmesi



Etkili Gözlemlerin Belirlenmesi

- Regresyon analizinin sonuçları ve sonuçların genellenebilirliği **birkaç gözlemle değişebilir.** Dolayısıyla bu gözlemlerin etkilerinin değerlendirilmesi için belirlenmesi gerekir.

- Etkili gözlem, gözlemlerdeki veya veri girişindeki bir hatadan kaynaklanabilir. Bu durumda birey analizden çıkarılabilir veya veri düzeltilebilir.

- Sıradışı bir durumla açıklanabilen, ender karşılaşılan geçerli bir gözlem analizden çıkarılabilir. Halbuki olası bir açıklaması olmayan, ender karşılaşılan bir gözlemi bir neden olmadan çıkarmak problemlidir ancak gözlemin analize dahil edilmesi de savunulamayabilir. **Bu durumda analizlerin gözlem dahil edilerek ve dahil edilmeyerek tekrarlanması önerilir.**

Etkili Gözlemlerin Belirlenmesi

Cook's D için kesme noktası 4/15= 0.267, 8. ve 11. gözlemler bu sınırı aşıyor

```
influence.measures(model, infl = influence(model))
```

```
## Influence measures of
##       lm(formula = Performans ~ Motivasyon + Kaygi, data = performans) :
##
##      dfb.1_  dfb.Mtvs dfb.Kayg   dffit cov.r   cook.d    hat inf
## 1   0.95795 -0.960377 -0.16108  1.1422 0.896 0.372039 0.3012
## 2  -0.25252  0.415588 -0.15745 -0.4984 1.422 0.084595 0.2501
## 3   0.14949 -0.263213  0.00861 -0.3676 1.209 0.045656 0.1388
## 4  -0.04850  0.110703 -0.01220  0.1872 1.332 0.012398 0.1026
## 5  -0.05212 -0.059328  0.11625 -0.1413 1.796 0.007232 0.2869   *
## 6  -0.50564 -0.103904  0.68481 -0.8203 0.924 0.201071 0.2200
## 7  -0.09262  0.115511 -0.02600 -0.1686 1.409 0.010150 0.1264
## 8   1.12071 -0.626060 -0.99660 -1.3623 1.570 0.574244 0.4905   *
## 9   0.07075 -0.094818  0.07387  0.3019 1.061 0.030173 0.0774
## 10 -0.00254  0.000079 -0.00142 -0.0191 1.390 0.000132 0.0670
## 11 -0.76153  1.357765 -0.15466  1.6459 0.228 0.510106 0.2092   *
## 12 -0.07535  0.006834  0.07308 -0.1166 1.424 0.004897 0.1142
## 13 -0.06170 -0.043739  0.14852  0.1670 1.965 0.010089 0.3487   *
## 14 -0.00051 -0.015642  0.01231 -0.0267 1.490 0.000260 0.1297
## 15 -0.07231  0.081208 -0.00672 -0.1135 1.472 0.004648 0.1373
```

Etkili Gözlemlerin Belirlenmesi

DFBETA için kesme noktası ise 2/(15^1/2) = 0.516 hat degerleri ise levarge a karşılık geliyor

```r
influence.measures(model, infl = influence(model))
```

```
## Influence measures of
##      lm(formula = Performans ~ Motivasyon + Kaygi, data = performans) :
##
##      dfb.1_  dfb.Mtvs dfb.Kayg   dffit cov.r   cook.d    hat inf
## 1   0.95795 -0.960377 -0.16108  1.1422 0.896 0.372039 0.3012
## 2  -0.25252  0.415588 -0.15745 -0.4984 1.422 0.084595 0.2501
## 3   0.14949 -0.263213  0.00861 -0.3676 1.209 0.045656 0.1388
## 4  -0.04850  0.110703 -0.01220  0.1872 1.332 0.012398 0.1026
## 5  -0.05212 -0.059328  0.11625 -0.1413 1.796 0.007232 0.2869   *
## 6  -0.50564 -0.103904  0.68481 -0.8203 0.924 0.201071 0.2200
## 7  -0.09262  0.115511 -0.02600 -0.1686 1.409 0.010150 0.1264
## 8   1.12071 -0.626060 -0.99660 -1.3623 1.570 0.574244 0.4905   *
## 9   0.07075 -0.094818  0.07387  0.3019 1.061 0.030173 0.0774
## 10 -0.00254  0.000079 -0.00142 -0.0191 1.390 0.000132 0.0670
## 11 -0.76153  1.357765 -0.15466  1.6459 0.228 0.510106 0.2092   *
## 12 -0.07535  0.006834  0.07308 -0.1166 1.424 0.004897 0.1142
## 13 -0.06170 -0.043739  0.14852  0.1670 1.965 0.010089 0.3487   *
## 14 -0.00051 -0.015642  0.01231 -0.0267 1.490 0.000260 0.1297
## 15 -0.07231  0.081208 -0.00672 -0.1135 1.472 0.004648 0.1373
```

teşekkürler