



Çoklu Regresyon



Bir Örnek

Dr. Kubra Atalay Kabasakal
Bahar 2023

Kategorik Bağımsız bir Değişken ile Çoklu Regresyon

- Regresyon modellerinde bağımsız bir değişken sürekli veya kategorik olabilir.
 - Bir regresyon analizine kategorik bir değişkeni dahil ederken, regresyon modelinin değişkenin düzeylerindeki farklılıkları doğru olarak kestirmesini sağlamak için, değişkenin düzeylerinin **yeniden kodlanması** gerekmektedir.
 - Kategorik değişkenleri kodlamanın en basit yöntemi **dummy (yapay)** kodlamadır.
 - Dummy kodlama ile kategorik değişkenin düzeylerine sayısal değerler atanarak dummy değişken(ler) oluşturulur.
- Dummy değişken kategorik bir değişkenin düzeylerinin sayısal gösterimidir.

Kategorik Bağımsız bir Değişken ile Çoklu Regresyon

- Bir bireyin k tane düzeye sahip kategorik bağımsız bir değişkenin belli bir düzeyine ilişkin üyeliğini temsil eden $k-1$ tane dummy değişken oluşturulur.
- Eğer bir birey kategorik değişkenin birinci düzeyindeyse birinci dummy değişkene 1 değeri verilir, birey değişkenin başka bir düzeyindeyse birinci dummy değişkene 0 değeri verilir.
- Eğer aynı birey kategorik değişkenin ikinci düzeyindeyse ikinci dummy değişkene 1 değeri verilir, birey değişkenin başka bir düzeyindeyse ikinci dummy değişkene 0 değeri verilir.
- Eğer aynı birey kategorik değişkenin $(k-1)$. düzeyindeyse $(k-1)$. dummy değişkene 1 değeri verilir, birey değişkenin başka bir düzeyindeyse $(k-1)$. dummy değişkene 0 değeri verilir.
- Böyle bir kodlama ile oluşturulan her bir dummy değişken iki düzeye sahiptir: 1 ve 0

Kategorik Bağımsız bir Değişken ile Çoklu Regresyon

- Bir kategorik değişken için oluşturulan dummy değişkenlerle gerçekleştirilen bir regresyon analizinde regresyon katsayılarının yorumlanması aşağıdaki gibidir:
- **Birinci dummy değişkene ilişkin katsayı**, diğer bütün değişkenler kontrol altına alındığında, **kategorik değişkenin birinci düzeyi ve kategorik değişkenin son düzeyi** arasındaki bağımlı değişkenin yordanan değerinin farkıdır.

Kategorik Bağımsız bir Değişken ile Çoklu Regresyon

- **İkinci dummy değişkene ilişkin katsayı**, diğer bütün değişkenler kontrol altına alındığında, **kategorik değişkenin ikinci düzeyi** ve **kategorik değişkenin son düzeyi** arasındaki bağımlı değişkenin yordanan değerinin farkıdır.
- **j. dummy değişkene ilişkin katsayı**, diğer bütün değişkenler kontrol altına alındığında, **kategorik değişkenin j. düzeyi** ve **kategorik değişkenin son düzeyi** arasındaki bağımlı değişkenin yordanan değerinin farkıdır.

Dummy Değişkenlerin Oluşturulması Örnek: Performansın Yordanması

- Üç düzeyi (evli, bekar ve diğer) bulunan medeni durum değişkeni için biri "Evli", diğeri "Bekar" olarak adlandırılan iki dummy değişken (örneğin, D1 ve D2) oluşturulabilir.

Medeni Durum	D1 (Evli)	D2 (Bekar)	D3 (Diğer)
Evli	1	0	0
Bekar	0	1	0
Diğer	0	0	1

- Referans Grup: Diğer
- Kategorik değişkenin **üç düzeyini** göstermek için üç gösterge değişkenine ihtiyaç yoktur. Değişkenin düzeyleri sadece **iki göstergeyle tanımlanmıştır**.
- D1 değişkeni için 0 değerine, D2 değişkeni için 0 değerine sahip bir birey diğer kategorisine aittir.

Dummy Değişkenlerin Oluşturulması Örnek: Performansın Yordanması

- D1 ve D2 bağımsız değişkenleri ile gerçekleştirilen bir regresyon analizinde kestirilen b_1 eğim katsayısı evliler ve diğerleri arasındaki yordanan matematik performansı farkını, b_2 eğim katsayısı ise bekarlar ve diğerleri arasındaki yordanan matematik performansı farkını belirtir.
- Evliler ve bekarlar arasındaki yordanan matematik performansı farkı, birinci ve ikinci medeni durum katsayıları arasındaki farktır: $b_1 - b_2$

Dummy Değişkenlerin Oluşturulması Örnek: Performansın Yordanması

```
library(haven)
Performansd1 <- read_sav("Performansd1.sav")
summary(Performansd1)
```

```
##      Performans      Motivasyon      Kaygi      Guven
## Min.      : 4.1    Min.      :22.0    Min.      :10.7    Min.      : 8.8
## 1st Qu.:11.9    1st Qu.:34.0    1st Qu.:15.5    1st Qu.:17.1
## Median :18.0    Median :40.0    Median :18.3    Median :22.0
## Mean    :18.2    Mean    :39.9    Mean    :18.1    Mean    :21.6
## 3rd Qu.:22.9    3rd Qu.:47.0    3rd Qu.:18.8    3rd Qu.:25.6
## Max.     :35.5    Max.     :55.0    Max.     :28.2    Max.     :38.7
##      Medeni
## Length:15
## Class :character
## Mode  :character
##
##
##
```


Dummy Değişkenlerin Oluşturulması Örnek: Performansın Yordanması

```
library(fastDummies)
# Performansd1$D1 <- ifelse(Performansd1$Medeni == "Evli", 1, 0)
# Performansd1$D2<- ifelse(Performansd1$Medeni == "Bekar", 1, 0)
dataf <- dummy_cols(Performansd1, select_columns = 'Medeni')
summary(dataf)
```

```
##      Performans      Motivasyon      Kaygi      Guven
## Min.      : 4.1    Min.      :22.0    Min.      :10.7    Min.      : 8.8
## 1st Qu.:11.9    1st Qu.:34.0    1st Qu.:15.5    1st Qu.:17.1
## Median :18.0    Median :40.0    Median :18.3    Median :22.0
## Mean     :18.2    Mean     :39.9    Mean     :18.1    Mean     :21.6
## 3rd Qu.:22.9    3rd Qu.:47.0    3rd Qu.:18.8    3rd Qu.:25.6
## Max.     :35.5    Max.     :55.0    Max.     :28.2    Max.     :38.7
##      Medeni      Medeni_1      Medeni_2      Medeni_3
## Length:15      Min.      :0.000    Min.      :0.000    Min.      :0.000
## Class :character 1st Qu.:0.000    1st Qu.:0.000    1st Qu.:0.000
## Mode  :character Median :0.000    Median :0.000    Median :0.000
##                Mean     :0.267    Mean     :0.467    Mean     :0.267
##                3rd Qu.:0.500    3rd Qu.:1.000    3rd Qu.:0.500
##                Max.      :1.000    Max.      :1.000    Max.      :1.000
```

Dummy Değişkenlerin Oluşturulması Örnek: Performansın Yordanması

```
model_dummy <- lm(Performans ~  
                  Medeni_1 + Medeni_2 ,  
                  data=datas)  
model_dummy
```

```
##  
## Call:  
## lm(formula = Performans ~ Medeni_1 + Medeni_2, data = datas)  
##  
## Coefficients:  
## (Intercept)      Medeni_1      Medeni_2  
##      19.644        -6.623         0.639
```

- Kesişim katsayısı b_0 19.644 matematik değerine eşittir. Bu değer, medeni durumu diğer olan öğrencilerin yordanan performans puanıdır.
- Evli için standartlaştırılmamış eğitim katsayısı b_1 -6.623 değerine eşittir. Bu değer, **evliler ve diğerleri arasındaki** yordanan matematik performans puanları farkının -6.623 birim olduğunu önerir.
- **$19.644 - 6.623 = 13.021$** medeni durumu evli olan öğrencilerin yordanan matematik performans puanıdır.

Dummy Değişkenlerin Oluşturulması Örnek: Performansın Yordanması

```
model_dummy
```

```
##  
## Call:  
## lm(formula = Performans ~ Medeni_1 + Medeni_2, data = dataf)  
##  
## Coefficients:  
## (Intercept)      Medeni_1      Medeni_2  
##      19.644      -6.623       0.639
```

- benzer şekilde, Bekar için standartlaştırılmamış eğitim katsayısı b_{21} 0.639 değerine eşittir. Bu değer, **bekarlar ve diğerleri arasındaki** yordanan matematik performans puanları farkının 0.639 birim olduğunu önerir.
- $19.644 + 0.639 = 20.283$ medeni durumu bekar olan öğrencilerin yordanan matematik performans puanıdır.

Çoklu Regresyonda İki-Yönlü Etkileşim Etkisi

- Eğer bir çoklu regresyon modelinde bir bağımsız değişken ile bağımlı değişken arasındaki ilişkinin büyüklüğü diğer bir bağımsız değişkenin düzeyine göre değişirse, etkileşim gözlenir.
- Etkileşim etkisi, düzenleyici (moderator) etki olarak da bilinmektedir.

Coklu Regresyonda Iki-Yönlü Etkileşim Etkisi

- Etkileşim terimi X_1 değişkeninin değerlerinin aracı X_2 değişkeninin değerleriyle çarpılmasıyla oluşan bileşik bir değişkendir.
- Regresyon eşitliği aşağıdaki gibidir:

$$Y = a + b_1x_{1i} + b_2x_{2i} + b_3x_{1i}x_{2i}$$

- Burada, b_3 katsayısı aracı etki olup X_2 değişkenin değeri değişirken X_1 değişkeninin etkisindeki birim değişimi belirtir.

Çoklu Regresyonda İki- Yönlü Etkileşim Etkisi

$$Y = a + b_1x_{1i} + b_2x_{2i} + b_3x_{1i}x_{2i}$$

- b_1 katsayısı X_2 değişkenine ilişkin değer sıfırken X_1 değişkeninin etkisini,
- b_2 katsayısı X_1 değişkenine ilişkin değer sıfırken X_2 değişkeninin etkisini belirtir (Etkileşim etkisinin bulunmadığı modelde, b_1 katsayısı X_2 değişkeninin bütün düzeylerinde X_1 değişkeninin etkisini, b_2 katsayısı X_1 değişkeninin bütün düzeylerinde X_2 değişkeninin etkisini temsil eder).

Çoklu Regresyonda İki- Yönlü Etkileşim Etkisi

- Bağımsız değişkenin toplam etkisini belirlemek için değişkenin ayrı ve aracı etkisi bir araya getirilmelidir.
- X_1 değişkeninin X_2 değişkeninin herhangi bir değerindeki toplam etkisi aşağıdaki eşitlikle hesaplanabilir:

$$b_{toplam} = b_1 + b_3 X_2$$

Çoklu Regresyonda İki- Yönlü Etkileşim Etkisi

-Sunumdaki örnek Wagner, Compas ve Howell (1988)'in çalışmasından gelmektedir.

- Wagner ve diğerleri çalışmalarında **daha fazla strese maruz kalan bireylerin psikolojik belirtileri daha yüksek düzeyde göstereceğini** önermiştir.
 - Ancak bir birey stresiyle basa çıkmasına yardımcı olacak yüksek düzeyde **sosyal desteğe sahipse**, belirtilerin stres arttıkça daha yavaş artmasının bekleneceğini, daha az sosyal desteğe sahip bireyler içinse, semptomların stres arttıkça daha hızlı artmasının bekleneceğini belirttiler.
-
- Wagner, B. M., Compas, B. E., & Howell, D. C. (1988). Daily and major life events: A test of an integrative model of psychosocial stress. American Journal of Community Psychology, 16 (2), 189-205.

Çoklu Regresyonda İki- Yönlü Etkileşim Etkisi

"Hassles.sav" SPSS veri dosyası bir ID değişkeni ve ID değişkeni dışında dört değişken ve 56 üniversite birinci sınıf öğrencisini içermektedir. Veri setine aşağıdaki web adresinden ulaşılabilir:

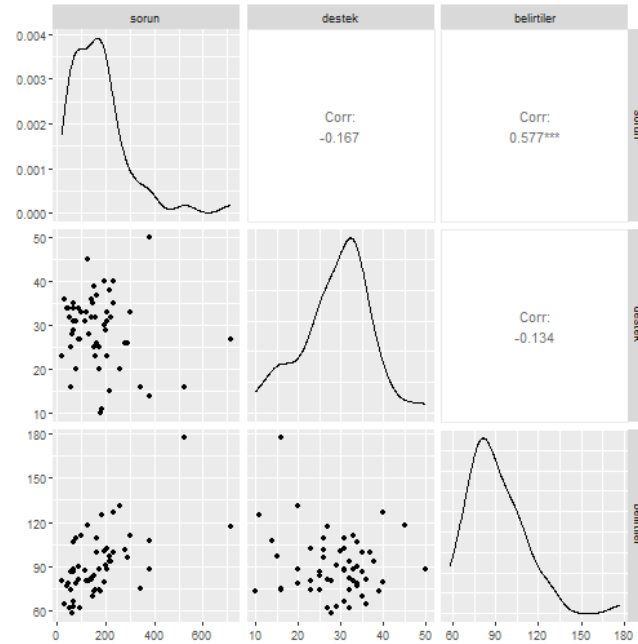
- <https://www.uvm.edu/~dhowell/methods8/DataFiles/DataSets.html>

```
library(haven)
zorluklar <- read_sav("Hassles.sav")[,c(1,2,4,5)]
colnames(zorluklar) <- c("id","sorun","destek","belirtiler")
```

Çoklu Regresyonda İki- Yönlü Etkileşim Etkisi

İlk olarak değişkenler arasındaki ilişkilere bakılır.

```
GGally::ggpairs(zorluklar[, -1])
```



Çoklu Regresyonda İki- Yönlü Etkileşim Etkisi

- Destek, sorun veya belirtiler ile ilişkili olmamasına rağmen, beklendiği gibi, Zorluklar ve Belirtiler arasında istatistiksel olarak bir ilişki vardır ($r = 0,577$).
- Ancak bu bulgular sorun ve belirtiler arasındaki ilişkisine destek düzeyine bağlı olup olmadığı sorusuna cevap vermemektedir

Çoklu Regresyonda İki- Yönlü Etkileşim Etkisi

- Sorun ve Destek etkileşimini test etmek için sorun değerleri ile destek değerlerinin çarpımıyla yeni bir değişken oluşturulur.
- Ancak bu iki değişkene ilişkin değerlerin çarpılmasıyla oluşan değişkenin analize dahil edilmesinde iki problem ortaya çıkacaktır.
 - Sorun veya Destek değişkenlerinden biri veya ikisi, çarpımlarıyla oluşan değişken ile yüksek düzeyde korelasyona sahip olacaktır ki bu da veride **çoklu bağlantı problemine** neden olacaktır.

Çoklu Regresyonda İki- Yönlü Etkileşim Etkisi

- Regresyon analizinde Sorun veya Destek değişkeninin herhangi bir etkisi diğer değişkenin değerinin 0 olduğu durumda değerlendirilecektir.
 - Diğer bir ifadeyle, Sorun üzerindeki test hiç sosyal desteği olmayan bir katılımcı için Sorunların belirtilerle ilişkili olup olmadığı testi olacaktır.
- Benzer şekilde, destek üzerindeki test hiç Sorunları olmayan katılımcılar için değerlendirilecektir.
- Hem **çoklu bağlantı problemi**, hem de **ana etkilerden birinin diğer ana etkinin uç değerinde değerlendirilmesi** problemi istenmeyen durumlardır.

Çoklu Regresyonda İki- Yönlü Etkileşim Etkisi

- Bahsedilen problemlerle basa çıkmak için **sorun değişkeni ve destek değişkeni** merkezlenebilir.
- Bunun için her bir değişkene ilişkin bireysel gözlemlerden ilgili değişkenin ortalaması çıkarılarak sapma puanları hesaplanacaktır.

```
zorluklar$csorun <- zorluklar$sorun -mean(zorluklar$sorun)
zorluklar$cdestek <- zorluklar$destek -mean(zorluklar$destek)
zorluklar$cross   <- zorluklar$sorun*zorluklar$destek
zorluklar$cross_m <- zorluklar$csorun*zorluklar$cdestek
```

Coklu Regresyonda Iki - Yönlü Etkileşim Etkisi

- Değişkenler merkezlendikten sonra merkezlenen sorun değişkeni için 0 değeri sorun değişkeninin ortalama düzeyindeki katılımcıları temsil ederken, merkezlenen destek değişkeni için 0 değeri destek değişkeninin ortalama düzeyindeki katılımcıları temsil eder.
- Böylece ana etkiler diğer değişkenin uygun düzeyinde değerlendirilir.

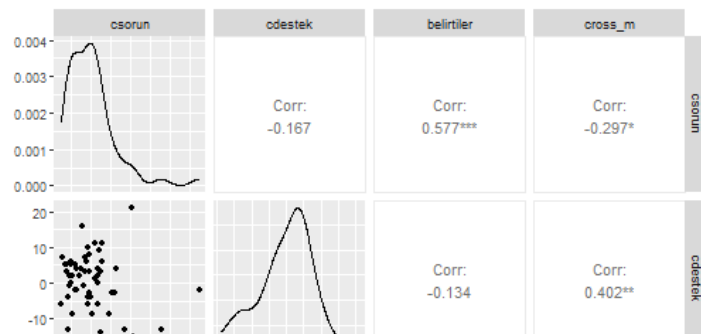
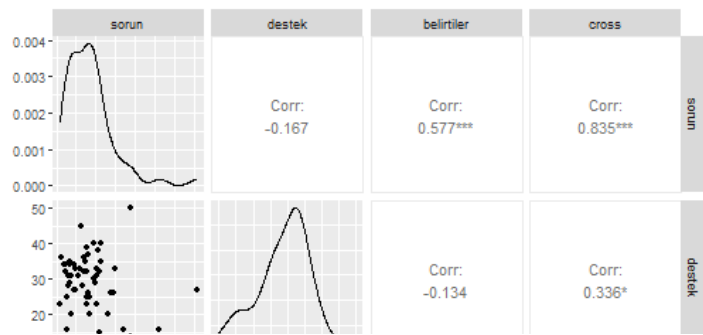
Çoklu Regresyonda İki Yönlü Etkileşim Etkisi

- Değişkenlerin merkezlenmesiyle sorun ve destek değişkenleri arasındaki **çoklu bağlantı** da önemli ölçüde düşecektir.
- Merkezlenen sorun, Merkezlenen destek, Merkezlenen sorun değişkeninin değerleri ile Merkezlenen Destek değişkeninin değerlerinin çarpılmasıyla elde edilen etkileşim terimi ve Belirtiler arasındaki korelasyon matrisi incelenir.

GGally::ggpairs(zorluklar[,c(2:4,7)])

GGally::ggpairs(zorluklar[,c(5,6,4,8)])

```
## plot: [1,1] [==>-----## plot: [1,1] I=6%-est: 0s-----
## plot: [1,2] [====>-----## plot: [1,2] I=12%-est: 0s-----
## plot: [1,3] [=====>-----## plot: [1,3] I=19%-est: 0s-----
## plot: [1,4] [=====>-----## plot: [1,4] I=25%-est: 0s-----
## plot: [2,1] [=====>-----## plot: [2,1] I=31%-est: 0s=>-----
## plot: [2,2] [=====>-----## plot: [2,2] I=38%-est: 0s====>-----
## plot: [2,3] [=====>-----## plot: [2,3] I=44%-est: 0s=====>-----
## plot: [2,4] [=====>-----## plot: [2,4] I=50%-est: 0s=====>-----
## plot: [3,1] [=====>-----## plot: [3,1] I=56%-est: 0s=====>-----
## plot: [3,2] [=====>-----## plot: [3,2] I=62%-est: 0s=====>-----
## plot: [3,3] [=====>-----## plot: [3,3] I=69%-est: 0s=====>-----
## plot: [3,4] [=====>-----## plot: [3,4] I=75%-est: 0s=====>-----
## plot: [4,1] [=====>-----## plot: [4,1] I=81%-est: 0s=====>-----
## plot: [4,2] [=====>-----## plot: [4,2] I=88%-est: 0s=====>-----
## plot: [4,3] [=====>-----## plot: [4,3] I=94%-est: 0s=====>-----
## plot: [4,4] [=====>-----## plot: [4,4] I=100%-est: 0s=====>-----
```



Çoklu Regresyonda İki Yönlü Etkileşim Etkisi

- Belirtilerin bağımlı değişken, merkezlenen sorunların ve merkezlenen Desteğin bağımsız değişken olarak modellendiği regresyonda, iki bağımsız değişkenin etkileşimini incelemek üzere etkileşim terimi de regresyon modeline eklenir.

```
cross_model <- lm(belirtiler ~ csorun + cdestek + cross_m ,data=zorluklar);broom::glance(cross_model)
```

```
## # 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    0.388    0.353  16.9    11.0 1.05e-5     3  -236.  481.  492.  14840.
## # ... with 2 more variables: df.residual <int>, nobs <int>, and abbreviated
## #   variable names 1adj.r.squared, 2statistic, 3deviance
```

- R^2 değeri yaklaşık 0.39 olup belirtilerdeki varyansın yaklaşık %39'unun doğrusal regresyon modeli tarafından açıklandığı anlamına gelmektedir.

Çoklu Regresyonda İki Yönlü Etkileşim Etkisi

```
broom::tidy(cross_model)
```

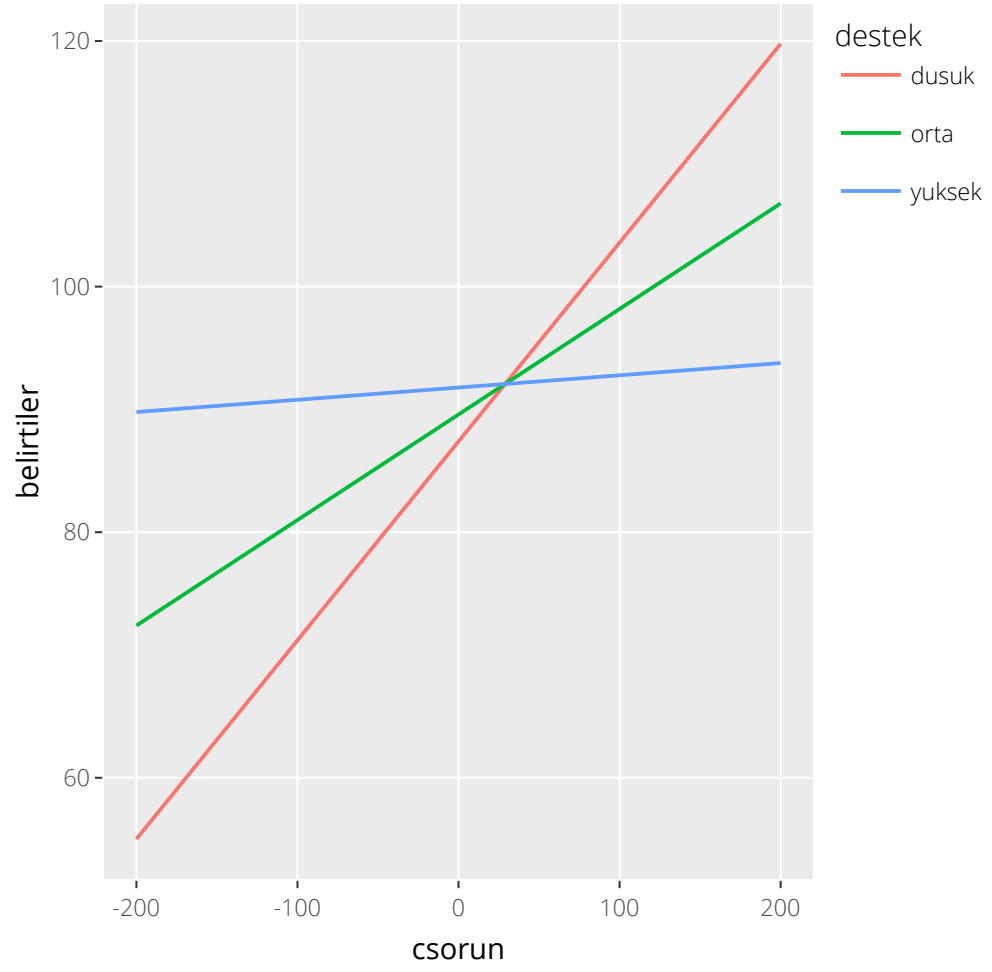
```
## # A tibble: 4 × 5
##   term          estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)  89.6         2.29      39.1    3.14e-40
## 2 csorun        0.0859      0.0192     4.47    4.22e- 5
## 3 cdestek       0.146       0.305     0.479    6.34e- 1
## 4 cross_m      -0.00506     0.00236    -2.14    3.68e- 2
```

- Hem merkezlenen sorun hem de etkileşim terimi istatistiksel olarak anlamlıdır (sırasıyla $p < 0.001$ ve $p = 0.037$). Ancak merkezlenen Destek istatistiksel olarak anlamlı değildir ($p = 0.634$). Destek değişkeni etkileşim teriminin hesaplanmasında yer aldığından, regresyon modelinde kalabilir.
- Etkileşim etkisinin anlamını yorumlamak için değişkenler arasındaki ilişkilerin grafik ile gösterimi yardımcı olabilir. En basit çözüm sosyal desteğin sabit düzeyleri için zorluklar ve psikolojik belirtiler arasındaki ilişkiye bakmaktır.

Çoklu Regresyonda İki Yönlü Etkileşim Etkisi

- Merkezlenen sosyal destek değişkeninin değerleri -21 ile +19 arasında değişmektedir. Bu değişken için düşük, orta ve yüksek değerleri temsil etmek üzere sırasıyla -15, 0 ve +15 değerleri seçilebilir.

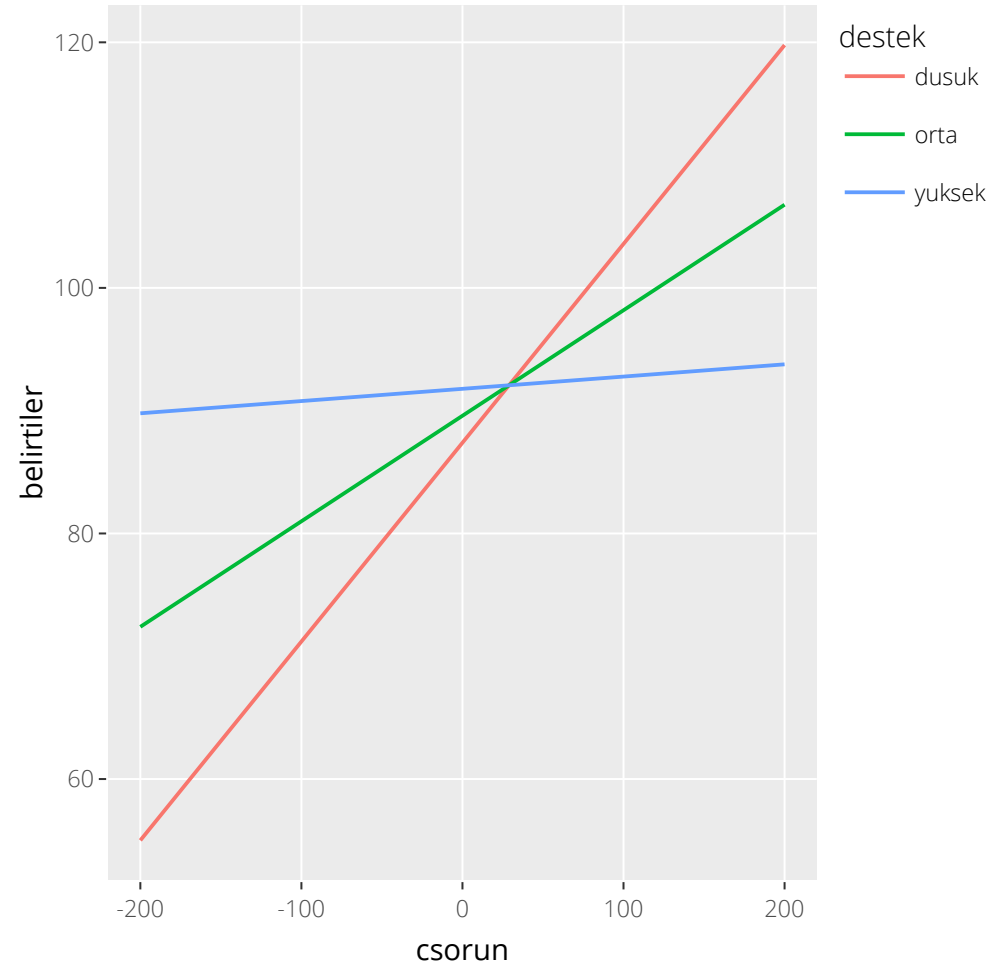
```
library(dplyr)
zorluklar <- zorluklar %>% mutate(
  cdestek_kat = case_when(
    cdestek <= -15 ~ "dusuk",
    cdestek >-15 & cdestek <15~ "orta",
    cdestek >=15 ~ "yuksek",
  )
)
zorluklar <- zorluklar %>% arrange(csorun)
```



Çoklu Regresyonda İki Yönlü Etkileşim Etkisi

- Yüksek düzeyde sosyal destek ile sorunlardaki artışlar psikolojik belirtilerde küçük artışlarla ilişkilendirilirler.
- Orta düzeyde sosyal destek ile sorunlardaki artışlar psikolojik belirtilerde daha büyük artışlarla ilişkilendirilir.
- Düşük düzeyde sosyal destek ile sorunlardaki artışlar psikolojik belirtilerde Dramatik artışlarla ilişkilendirilir.

ggplotly(p)



teşekkürler

