

Lojistik Regresyon Çözümlemesi

Hacettepe Üniversitesi öğrencilerinin bulundukları fakülte ve kitap okuma alışkanlıklarını etkileyen faktörler üzerine yapılan araştırmada veri seti; 20 nitelik ve 100 gözlemden oluşmaktadır. Araştırma içerisinde “Bölüm” kriteri dikkate alınmamıştır. Çünkü öğrencilerin bulunduğu bölümler anket üzerinden çoktan seçmeli olarak değil klasik bir şekilde alındığından bölüm kriteri araştırmaya dahil edilmemiştir.

```
1 ## veri girişi
2 ## veri seti içerisinde 20 nitelik bulunur.
3 ## 18 kategorik 1 sürekli ve 1 de tarih değişkeni bulunmaktadır.
4
5 verii <- kategorik_anket
6 str(veri)
```

Veri setinde bulunan 20 değişken içerisinde “Yaş” değişkeni dışından tüm değişkenler kategorik değişkendir. Bunun dışında “Zaman Damgası” değişkeni, tarih türünde bir veri olduğundan bu veri tipi de çalışmaya dahil edilmemiştir.

```
## tarih ve bölüm değişkeni analizden çıkartılır
## özellikle yetersiz gözlem olmasından kaynaklı olarak nitelik sayısı azaltılmalıdır
verii= verii[,-1]
verii= verii[,-4]
```

Gerekli olmayan değişkenlerin veri seti içerisinde çıkartılmasının ardından kategorik değişkenlerin analize dahil edilebilmesi için değişken dönüşümü yapılmalıdır. Özellikle multi (çok durumlu) kategorik değişkenler için dummy (kukla) değişken kullanılmalıdır. Bu nedenle lojistik çözümleme öncesi değişken dönüşümüne yer verilir.

```
## veri dönüşümü
## kategorik değişkenlerin one hot encoding dönüşümü
# 7 nitelik binary
# 10 nitelik multi

library(caret)
dmy <- dummyVars(" ~ .", data = verii)
yeni <- data.frame(predict(dmy, newdata = verii))
yeni
```

Lojistik regresyon çalışması için 18 farklı değişken kullandığımız için tüm değişkenleri analize dahil etmek, algoritmanın doğru bir sonuç vermesinin önüne geçebilir. Bu nedenle iki durumlu ve çok durumlu kategorik değişkenleri ayrı ayrı değerlendireceğiz.

Buradan Pr ($> |z|$) sonuçlarına bakıldığında anlamlılığa en yakın olan değişkenin “Üniversiteye başladıktan sonra kitap okuma alışkanlığında değişiklik oldu mu?” sorusudur. Buradaki evet yanıtının kitap okuma alışkanlığını artırdığı yönünde bir yorum yapılabilir. Ancak bu sonuç %100 olarak güvenilir bir sonuç değildir.

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.14922	2.48696	0.864	0.387
CinsiyetinizKadın	0.03736	0.44122	0.085	0.933
Yaşınız	-0.10009	0.10322	-0.970	0.332
X.En çok hangi eserleri okursunuz..Yerli	-0.04589	0.48271	-0.095	0.924
X.Çizgi.Romanların.kitap.okuma.alışkanlığı.kazandırdığını.düşünüyor.musunuz..Evet	0.54780	0.46965	1.166	0.243
X.Sevmediğiniz.kitapları.yarım.bırakır.mısınız..Evet	-0.64683	0.48518	-1.333	0.182
X.Okuduğunuz.bölümün.sizi.kitap.okumaya.teşvik.ettiğini.düşünüyor.musunuz..Evet	0.47676	0.46340	1.029	0.304
X.Üniversiteye.başladıktan.sonra.kitap.okuma.alışkanlığınızdadeğişiklik. oldu.mu..Evet	0.92141	0.47409	1.944	0.052
X.Çevreniz.kitap.okuma.alışkanlığınızı.etkiliyor.mu..Evet	-0.44478	0.47351	-0.939	0.348

Diğer değişkenlerin “Estimate” yani katsayılarına bakıldığında sıfıra oldukça yakın olduğunu görürüz. Bu durum da yine bu değişkenlerin araştırma üzerinde bir etkisinin olmadığını açıkça ifadesidir.

Ayrıca değişken anlamlılığının tespiti için de alternatif olarak Wald testinden faydalanılabilir.

H0: “...” Katsayı anlamlı değildir.

HS: “...” Katsayı anlamlıdır

Analysis of Deviance Table (Type II tests)

Response: X.Kitap.Okuma.Alışkanlığını.Nasıl.ölçeklendirirsiniz..Yüksek

	Df	Chisq	Pr(>Chisq)
CinsiyetinizKadın	1	0.0072	0.93252
Yaşınız	1	0.9404	0.33218
X.En çok hangi eserleri okursunuz..Yerli	1	0.0090	0.92427
X.Çizgi.Romanların.kitap.okuma.alışkanlığı.kazandırdığını.düşünüyor.musunuz..Evet	1	1.3605	0.24345
X.Sevmediğiniz.kitapları.yarım.bırakır.mısınız..Evet	1	1.7774	0.18247
X.Okuduğunuz.bölümün.sizi.kitap.okumaya.teşvik.ettiğini.düşünüyor.musunuz..Evet	1	1.0585	0.30356
X.Üniversiteye.başladıktan.sonra.kitap.okuma.alışkanlığınızdadeğişiklik. oldu.mu..Evet	1	3.7773	0.05195
X.Çevreniz.kitap.okuma.alışkanlığınızı.etkiliyor.mu..Evet	1	0.8824	0.34756

signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

%95 güvenle hiçbir katsayı anlamlı değildir. Ancak anlamlılığa en yakın katsayı;

“Üniversiteye başladıktan sonra kitap okuma alışkanlığında değişiklik oldu mu?” olarak bulunmuştur.

Yukarıdaki işlemlerle lojistik model kurulumunu gerçekleştirdik. Ancak kurulan model istatistiksel olarak anlamlı mı? Bunun tespit edilmesi oldukça önemlidir. Bunun için Hosmer Lemeshow testi yapılmalıdır.

H0: Lojistik regresyon modeline uyum vardır.

Hs: Lojistik regresyon modeline uyum yoktur.

Hosmer Lemeshow testinin sonuçlarına bakıldığında p-value = 0.9615'dir. %95 güven düzeyinde H0 hipotezi reddedilemez. Bunun sonucunda, Lojistik regresyon modeline uyum vardır yorumu yapılabilir.

Hosmer and Lemeshow goodness of fit (GOF) test

```
data: log_model$y, fitted(log_model)
X-squared = 2.505, df = 8, p-value = 0.9615
```

Aşağıdaki tablo ise, anlamlı olan lojistik regresyon modelinde yer alan katsayıların odds oranları ve güven aralığı yer almaktadır. Buradan “Üniversiteye başladıktan sonra kitap okuma alışkanlığında değişiklik oldu mu?” değişkenini anlamlı olarak kabul edersek, üniversite öğreniminin kitap okuma alışkanlığını 2.51 kat arttırdığı yorumu yapılabilir.

	OR	2.5 %	97.5 %
(Intercept)	8.5781654	0.0680705	1288.970458
CinsiyetinizKadın	1.0380667	0.4341324	2.471148
Yaşınız	0.9047518	0.7337555	1.105102
X.En çok.hangi.eserleri.okursunuz..Yerli	0.9551498	0.3714194	2.498524
X.Çizgi.Romanların.kitap.okuma.alışkanlığı.kazandırdığını.düşünüyor.musunuz..Evet	1.7294424	0.6875937	4.387262
X.Sevmediğiniz.kitapları.yarım.bırakır.mısınız..Evet	0.5237021	0.1971250	1.340719
X.okuduğunuz.bölümün.sizi.kitap.okumaya.teşvik.ettiğini.düşünüyor.musunuz..Evet	1.6108437	0.6517317	4.060140
X.Üniversiteye.başladıktan.sonra.kitap.okuma.alışkanlığınızda.değişiklik.öldü.mu..Evet	2.5128228	0.9997927	6.492502
X.Çevreniz.kitap.okuma.alışkanlığınızı.etkiliyor.mu..Evet	0.6409624	0.2460978	1.596581

Aşağıdaki sonuçlarda; fakülte, sınıf, genel akademik not ortalaması, okunulan kitapların seçiminde etkili olan faktörler ve en çok okunulan eser türleri ile yapılan lojistik modelin değerleri yer almaktadır (log_model2). Buradan Sınıf-3 katsayısının anlamlılığa en yakın değişken olduğu yorumu yapılabilir.

```
R 4.2.0 - C:/Users/Seher/Desktop/kategorik_ödev/
Coefficients: (3 not defined because of singularities)

              Estimate Std. Error z value Pr(>|z|)
(Intercept)      -18.33515  2751.69744   -0.007   0.9947
FakultenizEdebiyat -0.87803    1.13031   -0.777   0.4373
FakultenizFen     -1.05683    0.99313   -1.064   0.2873
FakultenizHukuk    0.50470    1.16531    0.433   0.6649
Fakultenizİktisadi.ve.Idari.Bilimler 0.39079    1.32295    0.295   0.7677
FakultenizMimarlık 15.95103  3956.18053    0.004   0.9968
FakultenizMühendislik -0.87710    1.09621   -0.800   0.4236
FakultenizSağlık  20.08189  3956.18071    0.005   0.9959
FakultenizSpor.Bilimleri 0.02896    1.18203    0.025   0.9805
Sınıfınız1        -0.16117    1.02135   -0.158   0.8746
Sınıfınız2        -0.65673    1.06084   -0.619   0.5359
Sınıfınız3        -1.53126    0.92863   -1.649   0.0992
Sınıfınız4        -1.23004    1.02592   -1.199   0.2305
Sınıfınız4..1      NA              NA          NA      NA
X.Genel.Akademik.Not.Ortalamanız.1.2 15.71181  2751.69734    0.006   0.9954
X.Genel.Akademik.Not.Ortalamanız.2.3 17.01070  2751.69730    0.006   0.9951
X.Genel.Akademik.Not.Ortalamanız.3.  20.54088  2751.69733    0.007   0.9940
X.Okuduğunuz.kitapların.seçiminde.etkili.olan.faktör.nedir..Kitabın.konusu 2.17524    1.38312    1.573   0.1158
X.Okuduğunuz.kitapların.seçiminde.etkili.olan.faktör.nedir..Tavsiye.üzerine 2.18251    1.53755    1.419   0.1558
X.Okuduğunuz.kitapların.seçiminde.etkili.olan.faktör.nedir..Yazar      2.11747    1.45795    1.452   0.1464
X.Okuduğunuz.kitapların.seçiminde.etkili.olan.faktör.nedir..Kitabın.sayfa.sayısı NA              NA          NA      NA
X.En.çok.hangi.tür.eserleri.okursunuz..Çizgi.Roman 2.44784    2.11136    1.159   0.2463
X.En.çok.hangi.tür.eserleri.okursunuz..Hikaye -0.10746    1.88260   -0.057   0.9545
X.En.çok.hangi.tür.eserleri.okursunuz..Roman    0.76425    1.82528    0.419   0.6754
X.En.çok.hangi.tür.eserleri.okursunuz..Şiir      NA              NA          NA      NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Oluşturulan 3. lojistik modele bakıldığında “Evinizde Ne kadar Kitap Vardır” sorusunun 1-50 cevabı istatistiksel olarak anlamlı çıkmıştır. (log_model3)

```
Coefficients: (4 not defined because of singularities)

              Estimate Std. Error z value Pr(>|z|)
(Intercept)      1.641179    1.225120    1.340   0.1804
X.Evinizde.ne.kadar.kitap.vardır..1.50 -1.395912    0.638060   -2.188   0.0287 *
X.Evinizde.ne.kadar.kitap.vardır..100.  0.003433    0.572105    0.006   0.9952
X.Evinizde.ne.kadar.kitap.vardır..51.100 NA              NA          NA      NA
X.Kitap.okumayı.en.sevdiğiniz.yer.neresidir..Ev -0.447231    0.802185   -0.558   0.5772
X.Kitap.okumayı.en.sevdiğiniz.yer.neresidir..Kafe -1.621297    1.326169   -1.223   0.2215
X.Kitap.okumayı.en.sevdiğiniz.yer.neresidir..Kütüphane 0.419566    1.020889    0.411   0.6811
X.Kitap.okumayı.en.sevdiğiniz.yer.neresidir..O.nun.yanı -1.111461    1.126415   -0.987   0.3238
X.Kitap.okumayı.en.sevdiğiniz.yer.neresidir..Park -0.851546    0.978350   -0.870   0.3841
X.Kitap.okumayı.en.sevdiğiniz.yer.neresidir..Toplu.Taşıma NA              NA          NA      NA
X.Sizi.okumaya.motive.eden.nedir..Bilgi.ihtiyacı 0.494562    0.706810    0.700   0.4841
X.Sizi.okumaya.motive.eden.nedir..Dinlenme.aktivitesi 0.256401    0.723781    0.354   0.7231
X.Sizi.okumaya.motive.eden.nedir..Okuma.alışkanlığı 1.271714    0.868648    1.464   0.1432
X.Sizi.okumaya.motive.eden.nedir..Üniversite.ödevleri NA              NA          NA      NA
X.Hangi.vakitler.kitap.okumayı.tercih.edersiniz..Akşam -1.015637    0.828034   -1.227   0.2200
X.Hangi.vakitler.kitap.okumayı.tercih.edersiniz..Gece -1.065427    0.833071   -1.279   0.2009
X.Hangi.vakitler.kitap.okumayı.tercih.edersiniz..Öğle -1.073462    0.916921   -1.171   0.2417
X.Hangi.vakitler.kitap.okumayı.tercih.edersiniz..Sabah NA              NA          NA      NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Python Statsmodels kütüphanesine baktığımızda ise “Evinizde Ne kadar Kitap Vardır? 1-50” değişkeninin haricinde, Sizi okumaya motive eden nedir sorusunda okuma alışkanlığı cevabı da istatistiksel olarak anlamlı çıkmıştır. (p-value= 0.041)

	coef	std err	z	P> z	[0.025	0.975]
const	-0.0343	0.199	-0.172	0.864	-0.425	0.357
Evinizde ne kadar kitap vardır?_1-50	-1.1971	0.494	-2.421	0.015	-2.166	-0.228
Evinizde ne kadar kitap vardır?_100+	0.5979	0.406	1.471	0.141	-0.199	1.395
Evinizde ne kadar kitap vardır?_51-100	0.5649	0.418	1.350	0.177	-0.255	1.385
Kitap okumayı en sevdiğiniz yer neresidir?_Ev	0.0278	0.471	0.059	0.953	-0.896	0.952
Kitap okumayı en sevdiğiniz yer neresidir?_Kafe	-0.9036	1.054	-0.857	0.392	-2.970	1.163
Kitap okumayı en sevdiğiniz yer neresidir?_Kütüphane	1.2946	0.780	1.660	0.097	-0.234	2.823
Kitap okumayı en sevdiğiniz yer neresidir?_O'nun yanı	-0.0895	0.968	-0.092	0.926	-1.987	1.808
Kitap okumayı en sevdiğiniz yer neresidir?_Park	-0.8573	0.820	-1.045	0.296	-2.465	0.750
Kitap okumayı en sevdiğiniz yer neresidir?_Toplu Taşıma	0.4937	0.746	0.662	0.508	-0.968	1.955
Sizi okumaya motive eden nedir?_Bilgi ihtiyacı	0.4044	0.456	0.886	0.375	-0.490	1.299
Sizi okumaya motive eden nedir?_Dinlenme aktivitesi	-0.4220	0.478	-0.883	0.377	-1.358	0.514
Sizi okumaya motive eden nedir?_Okuma alışkanlığı	1.2208	0.599	2.039	0.041	0.047	2.394
Sizi okumaya motive eden nedir?_Üniversite ödevleri	-1.2374	0.726	-1.704	0.088	-2.661	0.186
Hangi vakitler kitap okumayı tercih edersiniz?_Akşam	-0.6180	0.476	-1.298	0.194	-1.551	0.315
Hangi vakitler kitap okumayı tercih edersiniz?_Gece	-0.3773	0.486	-0.776	0.438	-1.331	0.576
Hangi vakitler kitap okumayı tercih edersiniz?_Sabah	1.0831	0.643	1.683	0.092	-0.178	2.344
Hangi vakitler kitap okumayı tercih edersiniz?_Öğle	-0.1221	0.625	-0.195	0.845	-1.347	1.103

İki Durumlu Kategorik Değişkenler ile Kurulan Lojistik Modelin Anlamlılığı

Log_model2 ve log_model3 isimli lojistik modellerin anlamlılıklarına bakalım. Bunun için yine Hosmer Lemeshow testlerini kullanacağız.

Çok durumlu kategorik değişkenler için kurulan log_model2 ve log_model_3 lojistik modelleri istatistiksel olarak doğru bir şekilde tanımlanmıştır diyebiliriz.

Çünkü;

H0: Lojistik regresyon modeline uyum vardır.

Hs: Lojistik regresyon modeline uyum yoktur

p-values > 0.05 olduğundan H0 reddedilemez ve kurulan iki modelin de Lojistik regresyon modeline uyumu olduğu söylenebilir.

```
> hoslem.test(log_model2$y, fitted(log_model2))

Hosmer and Lemeshow goodness of fit (GOF) test

data: log_model2$y, fitted(log_model2)
X-squared = 8.8087, df = 8, p-value = 0.3587

> hoslem.test(log_model3$y, fitted(log_model3))

Hosmer and Lemeshow goodness of fit (GOF) test

data: log_model3$y, fitted(log_model3)
X-squared = 6.8648, df = 8, p-value = 0.5513
```

Çok Durumlu Kategorik Değişkenlerin Odds Oranları ve Güven Aralıkları

```
> exp(cbind(OR = coef(log_model2), confint(log_model2)))  
waiting for profiling to be done...
```

	OR	2.5 %	97.5 %
(Intercept)	1.089294e-08	NA	1.250639e+168
FakültenizEdebiyat	4.156010e-01	4.087127e-02	3.730424e+00
FakültenizFen	3.475558e-01	4.528320e-02	2.363140e+00
FakültenizHukuk	1.656496e+00	1.704452e-01	1.798324e+01
Fakültenizİktisadi.ve.İdari.Bilimler	1.478150e+00	1.082496e-01	2.239257e+01
FakültenizMimarlık	8.461480e+06	0.000000e+00	NA
FakültenizMühendislik	4.159875e-01	4.327984e-02	3.417917e+00
FakültenizSağlık	5.265682e+08	0.000000e+00	NA
FakültenizSpor.Bilimleri	1.029387e+00	9.919658e-02	1.124879e+01
Sınıfınız1.	8.511498e-01	1.081390e-01	6.433856e+00
Sınıfınız2.	5.185431e-01	5.826842e-02	4.077854e+00
Sınıfınız3.	2.162629e-01	3.080516e-02	1.247715e+00
Sınıfınız4.	2.922805e-01	3.427360e-02	2.035961e+00
Sınıfınız4..1	NA	NA	NA
X.Genel.Akademik.Not.Ortalamanız.1.2	6.661201e+06	5.246272e-56	Inf
X.Genel.Akademik.Not.Ortalamanız.2.3	2.441469e+07	2.386665e-56	6.871354e+248
X.Genel.Akademik.Not.Ortalamanız.3.	8.332804e+08	3.059552e-52	Inf
X.Okuduğunuz.kitapların.seçiminde.etkili.olan.faktör.nedir..Kitabın.konusu	8.804294e+00	6.767926e-01	1.754388e+02
X.Okuduğunuz.kitapların.seçiminde.etkili.olan.faktör.nedir..Tavsiye.üzerine	8.868547e+00	5.285732e-01	2.453998e+02
X.Okuduğunuz.kitapların.seçiminde.etkili.olan.faktör.nedir..Yazar	8.310091e+00	5.445266e-01	1.867392e+02
X.Okuduğunuz.kitapların.seçiminde.etkili.olan.faktör.nedir..Kitabın.sayfa.sayısı	NA	NA	NA
X.En.çok.hangi.tür.eserleri.okursunuz..Çizgi.Roman	1.156334e+01	2.875018e-01	1.372678e+03
X.En.çok.hangi.tür.eserleri.okursunuz..Hikaye	8.981109e-01	2.310718e-02	4.594042e+01
X.En.çok.hangi.tür.eserleri.okursunuz..Roman	2.147382e+00	6.331891e-02	1.007676e+02
X.En.çok.hangi.tür.eserleri.okursunuz..Şiir	NA	NA	NA

There were 50 or more warnings (use warnings() to see the first 50)

```
> ## odds oranları ve güven aralıklarının elde edilmesi  
> exp(cbind(OR = coef(log_model3), confint(log_model3)))  
waiting for profiling to be done...
```

	OR	2.5 %	97.5 %
(Intercept)	5.1612525	0.51962538	68.2219916
X.Evinizde.ne.kadar.kitap.vardır..1.50	0.2476072	0.06712706	0.8379449
X.Evinizde.ne.kadar.kitap.vardır..100.	1.0034389	0.32024364	3.0798051
X.Evinizde.ne.kadar.kitap.vardır..51.100	NA	NA	NA
X.Kitap.okumayı.en.sevdiğiniz.yer.neresidir..Ev	0.6393962	0.11533842	2.9106787
X.Kitap.okumayı.en.sevdiğiniz.yer.neresidir..Kafe	0.1976422	0.01199313	2.4428178
X.Kitap.okumayı.en.sevdiğiniz.yer.neresidir..Kütüphane	1.5213011	0.19794592	12.1109126
X.Kitap.okumayı.en.sevdiğiniz.yer.neresidir..O.nun.yanı	0.3290778	0.03165107	2.8705578
X.Kitap.okumayı.en.sevdiğiniz.yer.neresidir..Park	0.4267548	0.05664328	2.8023182
X.Kitap.okumayı.en.sevdiğiniz.yer.neresidir..Toplu.Taşıma	NA	NA	NA
X.Sizi.okumaya.motive.eden.nedir..Bilgi.ihtiyacı	1.6397792	0.40855739	6.7607439
X.Sizi.okumaya.motive.eden.nedir..Dinlenme.aktivitesi	1.2922712	0.31044866	5.4951611
X.Sizi.okumaya.motive.eden.nedir..Okuma.alışkanlığı	3.5669616	0.67898786	21.4260049
X.Sizi.okumaya.motive.eden.nedir..Üniversite.ödevleri	NA	NA	NA
X.Hangi.vakitler.kitap.okumayı.tercih.edersiniz..Akşam	0.3621715	0.06141477	1.6971787
X.Hangi.vakitler.kitap.okumayı.tercih.edersiniz..Gece	0.3445807	0.05801937	1.6339700
X.Hangi.vakitler.kitap.okumayı.tercih.edersiniz..Öğle	0.3418230	0.05026904	1.9579181
X.Hangi.vakitler.kitap.okumayı.tercih.edersiniz..Sabah	NA	NA	NA

Geriye Doğru Seçim Yöntemi

Kurduğumuz lojistik modellerine dahil edilen tüm parametreler anlamlı değildir. Bu nedenle bizim en uygun parametrelerle bir model oluşturabilmemiz için, geriye doğru eleme yöntemi ile anlamsız parametrelerin dışarıda bırakıldığı bir yöntem kullanmaktır. Bu nedenle 3 lojistik modelimiz üzerinde de geriye doğru eleme yöntemini uygulayacağız. Bu yöntem aynı zamanda açıklayıcı değişkenler arasındaki bağlantı probleminin de önüne geçer.

Başlangıç AIC değeri: 141.1

```
> back_lr <- step(log_model)
Start: AIC=141.1
X. Kitap. Okuma. Alışkanlığınızı. Nasıl. Ölçeklendirirsiniz. .Yüksek ~
  cinsiyetinizkadın + Yaşınız + X. En. çok. hangi. eserleri. okursunuz. .Yerli +
  X. Çizgi. Romanların. kitap. okuma. alışkanlığı. kazandırdığını. düşünüyor. musunuz. .Evet +
  X. Sevmediğiniz. kitapları. yarım. bırakır. mısınız. .Evet +
  X. Okuduğunuz. bölümün. sizi. kitap. okumaya. teşvik. ettiğini. düşünüyor. musunuz. .Evet +
  X. Üniversiteye. başladıktan. sonra. kitap. okuma. alışkanlığınızda. değişiklik. oldu. mu. .Evet +
  X. Çevreniz. kitap. okuma. alışkanlığınızı. etkiliyor. mu. .Evet
```

Step: AIC=139.11

Step: AIC=137.12

Step: AIC=136.03

Step: AIC=135

Step: AIC=134.11

Step: AIC=133.78

En uygun lojistik model için en düşük AIC değerini ifade eden parametreleri tercih edeceğiz. Bu parametreler de “Sevmediğiniz.kitapları.yarım.bırakır.mısınız..Evet” ve “Üniversiteye.başladıktan.sonra.kitap.okuma.alışkanlığınızda.değişiklik.oldu.mu..Evet” parametreleridir.

	Df	Deviance	AIC
<none>		127.78	133.78
- X. Sevmediğiniz.kitapları.yarım.bırakır.mısınız..Evet	1	130.88	134.88
- X. Üniversiteye.başladıktan.sonra.kitap.okuma.alışkanlığınızda.değişiklik.oldu.mu..Evet	1	131.20	135.20

```
> |
```

Çok durumlu kategorik değişkenlerin ilk kısmını için oluşturduğum log_model2 lojistik modelinde en uygun parametreleri elde edebilmek için geriye doğru seçim yöntemini uygulayacağız.

```
> back_1f <- step(log_model2)
Start: AIC=132.42
X.Kitap.Okuma.Alışkanlığınızı.Nasıl.ölçeklendirirsiniz..Yüksek ~
FakültenizEdebiyat + FakültenizFen + FakültenizHukuk +
Fakültenizİktisadi.ve.Idari.Bilimler + FakültenizMimarlık +
FakültenizMühendislik + FakültenizSpor.Bilimleri +
Sınıfınız1. + Sınıfınız2. + Sınıfınız3. +
Sınıfınız3. + Sınıfınız4. + Sınıfınız4..1 +
X.Genel.Akademik.Not.Ortalamanız.1.2 + X.Genel.Akademik.Not.Ortalamanız.2.3 +
X.Genel.Akademik.Not.Ortalamanız.3. + X.Okuduğunuz.kitapların.seçiminde.etkili.olan.faktör.nedir..Kitabın.konusu +
X.Okuduğunuz.kitapların.seçiminde.etkili.olan.faktör.nedir..Tavsiye.üzerine +
X.Okuduğunuz.kitapların.seçiminde.etkili.olan.faktör.nedir..Yazar +
X.Okuduğunuz.kitapların.seçiminde.etkili.olan.faktör.nedir..Kitabın.sayfa.sayısı +
X.En.çok.hangi.tür.eserleri.okursunuz..Çizgi.Roman +
X.En.çok.hangi.tür.eserleri.okursunuz..Hikaye + X.En.çok.hangi.tür.eserleri.okursunuz..Roman +
X.En.çok.hangi.tür.eserleri.okursunuz..Şiir
```

Başlangıç AIC değeri: 132.42

Step: AIC=130.42
Step: AIC=128.44
Step: AIC=126.47
Step: AIC=124.52
Step: AIC=122.65
Step: AIC=120.88
Step: AIC=119.21
Step: AIC=117.91
Step: AIC=116.87
Step: AIC=115.11
Step: AIC=113.48
Step: AIC=112.27
Step: AIC=110.87
Step: AIC=109.91
Step: AIC=109.09
Step: AIC=109.08

En düşük AIC değeri 109.06' dır. Bizim log_model2 lojistik modelimiz için oluşturacağımız en iyi modelde aşağıdaki parametreler yer almalıdır.

```
Step: AIC=109.06
X.Kitap.Okuma.Alışkanlığınızı.Nasıl.ölçeklendirirsiniz..Yüksek ~
Sınıfınız3. + X.Genel.Akademik.Not.Ortalamanız.2.3 +
X.Genel.Akademik.Not.Ortalamanız.3.
```

	Df	Deviance	AIC
<none>		101.06	109.06
- X.Genel.Akademik.Not.Ortalamanız.2.3	1	104.20	110.20
- Sınıfınız3.	1	108.08	114.08
- X.Genel.Akademik.Not.Ortalamanız.3.	1	126.40	132.40

```
> |
```

Son olarak log_model3 lojistik modelimiz üzerine geriye doğru seçim yöntemini uygulayalım.


```

> back_lr <- step(log_model13)
Start: AIC=140.15
X.Kitap.okuma.Alışkanlığınızı.Nasıl.ölçeklendirirsiniz..Yüksek ~
  X.Evinizde.ne.kadar.kitap.vardır..1.50 + X.Evinizde.ne.kadar.kitap.vardır..100. +
  X.Evinizde.ne.kadar.kitap.vardır..51.100 + X.Kitap.okumayı.en.sevdiğiniz.yer.neresidir..Ev +
  X.Kitap.okumayı.en.sevdiğiniz.yer.neresidir..Kafe +
  X.Kitap.okumayı.en.sevdiğiniz.yer.neresidir..Kütüphane +
  X.Kitap.okumayı.en.sevdiğiniz.yer.neresidir..O.nun.yanı +
  X.Kitap.okumayı.en.sevdiğiniz.yer.neresidir..Park +
  X.Kitap.okumayı.en.sevdiğiniz.yer.neresidir..Toplu.Taşıma +
  X.Sizi.okumaya.motive.eden.nedir..Bilgi.ihhtiyacı + X.Sizi.okumaya.motive.eden.nedir..Dinlenme.aktivitesi +
  X.Sizi.okumaya.motive.eden.nedir..Okuma.alışkanlığı +
  X.Sizi.okumaya.motive.eden.nedir..Üniversite.ödevleri +
  X.Hangi.vakitler.kitap.okumayı.tercih.edersiniz..Akşam +
  X.Hangi.vakitler.kitap.okumayı.tercih.edersiniz..Gece +
  X.Hangi.vakitler.kitap.okumayı.tercih.edersiniz..Öğle +
  X.Hangi.vakitler.kitap.okumayı.tercih.edersiniz..Sabah

```

Başlangıç AIC değeri: 140.15

Step: AIC=138.15

Step: AIC=136.27

Step: AIC=134.45

Step: AIC=132.79

Step: AIC=132.2

Step: AIC=131.36

Step: AIC=130.62

Step: AIC=129.91

En küçük AIC değeri 129.9 olarak bulunmuştur. Bu değere karşılık gelen model, bizim için en iyi modeldir.

```

Step: AIC=129.9
X.Kitap.okuma.Alışkanlığınızı.Nasıl.ölçeklendirirsiniz..Yüksek ~
  X.Evinizde.ne.kadar.kitap.vardır..1.50 + X.Hangi.vakitler.kitap.okumayı.tercih.edersiniz..Akşam +
  X.Hangi.vakitler.kitap.okumayı.tercih.edersiniz..Gece +
  X.Hangi.vakitler.kitap.okumayı.tercih.edersiniz..Öğle

```

	Df	Deviance	AIC
<none>		119.90	129.90
- X.Hangi.vakitler.kitap.okumayı.tercih.edersiniz..Akşam	1	122.09	130.09
- X.Hangi.vakitler.kitap.okumayı.tercih.edersiniz..Gece	1	122.45	130.45
- X.Hangi.vakitler.kitap.okumayı.tercih.edersiniz..Öğle	1	123.65	131.65
- X.Evinizde.ne.kadar.kitap.vardır..1.50	1	130.55	138.55

Uygun Parametreler ile Yeni Lojistik Regresyon Modelinin Kurulması

En düşük AIC değerlerine karşılık gelen parametreleri seçerek yeniden lojistik model kuralım.

Parametrelerin anlamlılıklarına baktığımızda 3.sınıf cevabının en yüksek anlamlılığa sahip olan parametre olduğu söylenebilir. Bunun dışında Genel akademik ortalamanın 3+ olması ve Evinizde ne kadar kitap vardır sorusunun 1-50 cevabı anlamlı çıkan diğer parametrelerdir.

```
coefficients:
(Intercept) -0.26919 1.36011 -0.198 0.843112
X.Sevmediğiniz.kitapları.yarım.bırakır.mısınız..Evet -0.55403 0.60937 -0.909 0.363256
X.Üniversiteye.başladıktan.sonra.kitap.okuma.alışkanlığınızda.değişiklik.ordu.mu..Evet 1.03338 0.63330 1.632 0.102736
X.Genel.Akademik.Not.Ortalamanız.2.3 1.05262 0.72541 1.451 0.146764
Sınıfınız3. -1.53899 0.60109 -2.560 0.010457 *
X.Genel.Akademik.Not.Ortalamanız.3. 4.58959 1.30839 3.508 0.000452 ***
X.Evinizde.ne.kadar.kitap.vardır..1.50 -1.53019 0.69855 -2.191 0.028487 *
X.Hangi.vakitler.kitap.okumayı.tercih.edersiniz..Akşam -0.02218 1.16199 -0.019 0.984773
X.Hangi.vakitler.kitap.okumayı.tercih.edersiniz..Gece -0.15194 1.10884 -0.137 0.891010
X.Hangi.vakitler.kitap.okumayı.tercih.edersiniz..Öğle -0.84597 1.16057 -0.729 0.466044
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Wald test istatistiğinde çıkan sonuçlarda da yukarıda belirtilen parametreler anlamlı çıkmıştır.

```
Df    Chisq Pr(>Chisq)
1    0.8266 0.3632559
1    2.6625 0.1027363
1    2.1056 0.1467644
1    6.5553 0.0104573 *
1   12.3048 0.0004518 ***
1    4.7984 0.0284867 *
1    0.0004 0.9847725
1    0.0188 0.8910097
1    0.5313 0.4660436
```

Hoslem Testi ile uygun modelin anlamlı olup olmadığına baktığımızda ise p-value sonucunun %66 değerinde olması uygun modelimizin lojistik regresyon bakımından anlamlı olduğunu ortaya çıkarır.

```
> hoslem.test(uygun_model$y, fitted(uygun_model))
```

Hosmer and Lemeshow goodness of fit (GOF) test

```
data: uygun_model$y, fitted(uygun_model)
x-squared = 5.8714, df = 8, p-value = 0.6616
```

Uygun modele ilişkin odds oranları ve güven aralıkları:

Buradan özellikle genel akademik not ortalaması 3 ve üzeri olan öğrencilerin kitap okuma alışkanlığını 98 kat arttırdığı yorumu yapılabilir.

*** Odds oranlarının sağlıklı bir şekilde yorumlanabilmesi için ise güven aralıklarının 1 değerini içermemesi gerekir.

```
> ## odds oranları ve güven aralıklarının elde edilmesi
> exp(cbind(OR = coef(uygun_model), confint(uygun_model)))
waiting for profiling to be done...
```

	OR	2.5 %	97.5 %
(Intercept)	0.7640019	0.05495731	12.4287757
X. Sevmediğiniz kitapları yarım bırakır mısınız..Evet	0.5746319	0.16803133	1.8817301
X. Üniversiteye başladıktan sonra kitap okuma alışkanlığınızda değişiklik oldu.mu..Evet	2.8105433	0.83520797	10.3785614
X. Genel Akademik Not Ortalamanız 2.3 Sınıfınız.	2.8651387	0.71421144	12.8151980
X. Genel Akademik Not Ortalamanız 3.	0.2145981	0.06176460	0.6685146
X. Evinizde ne kadar kitap vardır..1. 50	98.4539540	10.69728786	2498.2597144
X. Hangi vakitler kitap okumayı tercih edersiniz..Akşam	0.2164947	0.04930686	0.8049412
X. Hangi vakitler kitap okumayı tercih edersiniz..Gece	0.9780664	0.08448120	9.0980256
X. Hangi vakitler kitap okumayı tercih edersiniz..Öğle	0.8590394	0.07937645	6.9032945
X. Hangi vakitler kitap okumayı tercih edersiniz..Diğer	0.4291391	0.03640192	3.8627965

Sonuç Olarak

Hacettepe Üniversitesi öğrencileri ile yapılan bu araştırmada ankete 7 farklı fakülte dahil edilmiştir. Lojistik regresyon ve ilişki testleri sonuçlarına göre Fakülte değişkeninin kitap okuma alışkanlığı üzerinde anlamlı bir etkisi bulunmamaktadır. Araştırma hakkında daha detaylı sonuçlar elde edilebilmesi için ise fakültelerden alınan gözlemlerin arttırılması tavsiye edilir. Bunun yanı sıra fakülte değişkeni anlamlı olarak çıkmasa da ankete katılan öğrencilerin belirttiği fikirlere göre üniversite öğreniminin kitap okuma alışkanlığı üzerinde pozitif yönlü bir etkisi vardır. Ayrıca öğrencilerin genel akademik not ortalamaları ile kitap okuma alışkanlıkları arasında da anlamlı bir ilişki bulunmuştur. Buradan öğrenci başarısı ile kitap okuma alışkanlığı arasındaki pozitif yönlü arttırıcı bir ilişki olduğundan bahsedilebilir. Son olarak özellikle odds oranlarının daha güvenilir bir şekilde yorumlanabilmesi için tüm fakültelerden alınan gözlemlerin arttırılması tavsiye edilir. Alınan 100 gözlem ile 7 farklı fakülte hakkında yorumda bulunabilmenin yeterli olmadığına kanaat edilmiştir.