

BÖLÜM 5

VAKA ÇALIŞMASI:
BEŞ BAĞIMSIZ DEĞİŞKEN

GİRİŞ

- Artık sıradan regresyon analizinin temel adımı korunmuş oldu.
- Bu bölümde bu yöntemin uygulaması gösterilecektir.
- Hesaplamalar ve regresyon sonuçlarının yorumları belirtilecektir.

KORKU BURNU HALİCİNDE SPARTİNA BİYOKÜTLE ÜRETİMİ

- İlgili veri North Carolina State Üniversitesinden Dr. Rick Linthurst (1979)'un doktora tezinin bir parçasından alınmıştır.
- Bu çalışmanın amacı Kuzey Carolina Korku Burnu Halicinde göl otu olan Spartina alterniflor'nın havaya ilişkin biyokütle üretimini etkileyen önemli toprak özelliklerini belirlemektir.

TAM MODEL İÇİN REGRESYON ANALİZİ

Başlangıç modelimiz *BIOMASS* olsun, Y , beş farklı bağımsız değişken artı kesişim katsayısı ile doğrusal ilişkili olarak karakterize edilsin. Böylece doğrusal model

$$Y = X\beta + \epsilon \quad (5.1)$$

Y , X ve β tanımlanarak belirlenmekte ve rassal hataların ϵ dağılımı hakkında uygun varsayımlar belirlenmektedir. Y , *BIOMASS*'in vektörü olup değerleri:

$$Y = (676 \quad 516 \quad \cdots \quad 1,560)'$$

TABLO 5.1. *Kuzey Carolina Korku Burnu Halicindeki hava biyokütle (BIO) ve alt katmanların beş fizyokimyasal özellikleri (SAL, pH, K, Na ve Zn). (Veri Dr. R. A. Linthurst'in izni ile kullanılmıştır.)*

<i>Obs.</i>	<i>Loc.</i>	<i>Type</i>	<i>BIO</i>	<i>SAL</i>	<i>pH</i>	<i>K</i>	<i>Na</i>	<i>Zn</i>
1	OI	DVEG	676	33	5.00	1,441.67	35,185.5	16.4524
2	OI	DVEG	516	35	4.75	1,299.19	28,170.4	13.9852
3	OI	DVEG	1,052	32	4.20	1,154.27	26,455.0	15.3276
4	OI	DVEG	868	30	4.40	1,045.15	25,072.9	17.3128
5	OI	DVEG	1,008	33	5.55	521.62	31,664.2	22.3312
6	OI	SHRT	436	33	5.05	1,273.02	25,491.7	12.2778
7	OI	SHRT	544	36	4.25	1,346.35	20,877.3	17.8225
8	OI	SHRT	680	30	4.45	1,253.88	25,621.3	14.3516
9	OI	SHRT	640	38	4.75	1,242.65	27,587.3	13.6826
10	OI	SHRT	492	30	4.60	1,281.95	26,511.7	11.7566
11	OI	TALL	984	30	4.10	553.69	7,886.5	9.8820
12	OI	TALL	1,400	37	3.45	494.74	14,596.0	16.6752
13	OI	TALL	1,276	33	3.45	525.97	9,826.8	12.3730
14	OI	TALL	1,736	36	4.10	571.14	11,978.4	9.4058
15	OI	TALL	1,004	30	3.50	408.64	10,368.6	14.9302
16	SI	DVEG	396	30	3.25	646.65	17,307.4	31.2865
17	SI	DVEG	352	27	3.35	514.03	12,822.0	30.1652
18	SI	DVEG	328	29	3.20	350.73	8,582.6	28.5901
19	SI	DVEG	392	34	3.35	496.29	12,369.5	19.8795
20	SI	DVEG	236	36	3.30	580.92	14,731.9	18.5056
21	SI	SHRT	392	30	3.25	535.82	15,060.6	22.1344
22	SI	SHRT	268	28	3.25	490.34	11,056.3	28.6101
23	SI	SHRT	252	31	3.20	552.39	8,118.9	23.1908
24	SI	SHRT	236	31	3.20	661.32	13,009.5	24.6917
25	SI	SHRT	340	35	3.35	672.15	15,003.7	22.6758
26	SI	TALL	2,436	29	7.10	528.65	10,225.0	0.3729
27	SI	TALL	2,216	35	7.35	563.13	8,024.2	0.2703
28	SI	TALL	2,096	35	7.45	497.96	10,393.0	0.3205
29	SI	TALL	1,660	30	7.45	458.38	8,711.6	0.2648
30	SI	TALL	2,272	30	7.40	498.25	10,239.6	0.2105

31	SM	DVEG	824	26	4.85	936.26	20,436.0	18.9875
32	SM	DVEG	1,196	29	4.60	894.79	12,519.9	20.9687
33	SM	DVEG	1,960	25	5.20	941.36	18,979.0	23.9841
34	SM	DVEG	2,080	26	4.75	1,038.79	22,986.1	19.9727
35	SM	DVEG	1,764	26	5.20	898.05	11,704.5	21.3864
36	SM	SHRT	412	25	4.55	989.87	17,721.0	23.7063
37	SM	SHRT	416	26	3.95	951.28	16,485.2	30.5589
38	SM	SHRT	504	26	3.70	939.83	17,101.3	26.8415
39	SM	SHRT	492	27	3.75	925.42	17,849.0	27.7292
40	SM	SHRT	636	27	4.15	954.11	16,949.6	21.5699
41	SM	TALL	1,756	24	5.60	720.72	11,344.6	19.6531
42	SM	TALL	1,232	27	5.35	782.09	14,752.4	20.3295
43	SM	TALL	1,400	26	5.50	773.30	13,649.8	19.5880
44	SM	TALL	1,620	28	5.50	829.26	14,533.0	20.1328
45	SM	TALL	1,560	28	5.40	856.96	16,892.2	19.2420

KORELASYON MATRİSİ

Çoklu regresyon analizine başlamanın en iyi yolu bağımlı değişkeni de içeren tüm değişkenler arasındaki korelasyon matrisini hesaplamaktır. Bu, değişkenler arasındaki basit doğrusal ilişkilere ilk bakışı sağlar. Korelasyon matrisi:

$$\hat{\rho} = S [W'(I - J/n)W] S, \quad (5.4)$$

ile elde edilir. Burada $n = 45$, I (45×45)'lik birim matris, J (45×45)'lik 1'ler matrisi, W (45×6)'lık BIOMASS (Y)'in ve beş bağımsız değişkenin matrisi ve S , her bir değişkenin düzeltilmiş kareler toplamının kareköklerinin karşılıklı köşegen matristir. Düzeltilmiş kareler toplamını $W'(I - J/n)W$ 'nin köşegen elemanları verir. Linthurst verisi için,

$$\hat{\rho} = \begin{bmatrix} Y & SAL & pH & K & Na & Zn \\ 1 & -.103 & .774 & -.205 & -.272 & -.624 \\ -.103 & 1 & -.051 & -.021 & .162 & -.421 \\ .774 & -.051 & 1 & .019 & -.038 & -.722 \\ -.205 & -.021 & .019 & 1 & .792 & .074 \\ -.272 & .162 & -.038 & .792 & 1 & .117 \\ -.624 & -.421 & -.722 & .074 & .117 & 1 \end{bmatrix}.$$

$\hat{\rho}$ matrisinin ilk satırını bağımsız değişkenlerin her birinin bağımlı değişken ile basit korelasyonlarını göstermektedir. İki değişken *pH* ve *Zn*, *BIOMASS* ile oldukça yüksek korelasyona sahiptir. Eğer regresyonda ayrı ayrı bağımsız değişken olarak kullanılırsa *BIOMASS*'daki değişmelerin *pH* ve *Zn* sırasıyla % 60 ($r^2 = .774^2$) ve % 39'na sebep olur.

TABLO 5.2. Beş bağımsız değişken SAL, pH, K, Na ve Zn'in BIOMASS üzerine regresyon sonuçları.

<i>Değişken</i>	$\hat{\beta}_j$	$s(\hat{\beta}_j)$	<i>t</i>	<i>Kısmi SS</i>
<i>SAL</i>	-30.285	24.031	-1.26	251,921
<i>pH</i>	305.525	87.879	3.48	1,917,306
<i>K</i>	-.2851	.3484	-.82	106,211
<i>Na</i>	-.0087	.0159	-.54	47,011
<i>Zn</i>	-20.676	15.055	-1.37	299,209

BIOMASS için Varyans Analizi

<i>Kaynak</i>	<i>d.f.</i>	<i>Kareler Toplamı</i>	<i>Ortalama Kare</i>
<i>Toplam</i>	44	19,170,963	
<i>Regresyon</i>		,984,700	2,596,940 $F = 16.37$
<i>Artık</i>	39	6,186,263	158,622

ÇOKLU REGRESYON SONUÇLARI: TAM MODEL

Beş bağımsız değişkenin hepsinin kullanıldığı analiz sonuçları Tablo 5.2'de verilmiştir. *BIOMASS* ile bağımsız değişkenler arasında yüksek ilişki vardır. Belirleme katsayısı $R^2 = 0.677$ 'dir. (belirleme katsayısı tanımını için Tablo 1.5, s. 15'e bakınız). Böylece *BIOMASS* kareler toplamının % 68'i beş bağımsız değişkendeki değişme ile ilişkilidir. Tüm beş regresyon katsayısının 0' a eşit bileşik sıfır hipotezi oldukça anlamlıdır; $F = 16.37$, $F_{(.01,5,39)} = 3.53$ ile karşılaştırıldığında.

TABLO 5.3. Etkileşimli matris program dili SAS/IML için yazılan regresyon analizi matris cebiri Y ve X 'in matris programında tanımladığı ve X 'in tam ranka sahip olduğu varsayılırsa

<i>SAS/IML program adımları^a</i>	<i>Hesaplanan matris</i>
$INVX=INV(X^*X);$ X^* ; X 'in devriğini gösterir	$(X'X)^{-1}$
$B=INVX*X^*Y;$	$\hat{\beta}$
$CF=SUM(Y)##2/NROW(X);$ $##2$ kareler $SUM(Y)$	$Y'(J/n)Y = (\sum Y)^2/n$
$SST=Y^*Y-CF;$ <i>BIOMASS</i> için Düzeltilmiş Kareler Toplamı	$Y'(I - J/n)Y$
$SSR=B^*X^*Y-CF;$ P 'nin hesaplanmasına gerek yok	$Y'(P - J/n)Y = SS(Regr)$
$SSE=SST-SSR;$	$Y'(I - P)Y = SS(Res)$

$$S2=SSE/(NROW(X)-NCOL(X)); \quad s^2$$

σ^2 tahminin serbestlik derecesi
= $n - r(\mathbf{X})$

$$SEB=SQRT(VECDIAG(INVX)*S2); \quad \hat{\beta} \text{ standart hatası}$$

“VECDIAG” köşegen vektörden
elde edilir.

$$T=B/SEB; \quad H_0 : \beta_j = 0 \text{ için } t$$

“/” eleman bölümünü göstermek-
tedir

$$PART=B\#\#2/VECDIAG(INVX); \quad \text{Kısmi Kareler Toplamı}$$

$$YHAT=X*B; \quad \hat{Y} \text{ için ortalama tahmin}$$

$$E=Y-YHAT; \quad \text{Tahmin edilen artıklar}$$

TABLO 5.4. SALINITY, pH, K ve Zn dört bağımsız değişkenin BIOMASS üzerine regresyon sonuçları

<i>Değişken</i>	$\hat{\beta}_j$	$s(\hat{\beta}_j)$	<i>t</i>	<i>Kısmi SS</i>
<i>Sal</i>	-35.94	21.48	-1.67	436,496
<i>pH</i>	293.9	84.5	3.48	1,885,805
<i>K</i>	-0.439	0.202	-2.17	732,606
<i>Zn</i>	-23.45	14.04	-1.67	434,796

Varyans Analizi

<i>Kaynak</i>	<i>d.f.</i>	<i>Kareler Toplamı</i>	<i>Ortalama Kare</i>
Toplam	44	19,170,963	
Regresyon	4	12,937,689	3,234,422
Kalıntı	40	6,233,274	155,832

TABLO 5.5. SALINITY, pH ve K üç bağımsız değişkenin BIOMASS üzerine regresyon sonuçları

<i>Değişken</i>	$\hat{\beta}_j$	$s(\hat{\beta}_j)$	t	<i>Kısmi SS</i>
<i>SAL</i>	-12.06	16.37	-.74	88,239
<i>pH</i>	410.21	48.83	8.40	11,478,835
<i>K</i>	-.490	.204	-2.40	935,178

Varyans Analizi

<i>Kaynak</i>	<i>d.f.</i>	<i>Kareler Toplam</i>	<i>Ortalama Kare</i>
<i>Toplam</i>	44	19,170,963	
<i>Regresyon</i>	3	12,502,893	4,167,631
<i>Kalıntı</i>	41	6,668,070	162,636

TABLO 5.6. *pH* ve *K* iki bağımsız değişkenin *BIOMASS* üzerine regresyon sonuçları

<i>Değişken</i>	$\hat{\beta}_j$	$s(\hat{\beta}_j)$	<i>t</i>	<i>Kısmi SS</i>
<i>pH</i>	412.04	48.50	8.50	11,611,782
<i>K</i>	-0.487	0.203	-2.40	924,266

Varyans Aalizi

<i>Kaynak</i>	<i>d.f.</i>	<i>Kareler Toplamı</i>	<i>Ortalama Kare</i>
Toplam	44	19,170,963	
Regresyon	2	12,414,654	6,207,327
Kalıntı	42	6,756,309	160,865

SON MODELİN SONUÇLARI

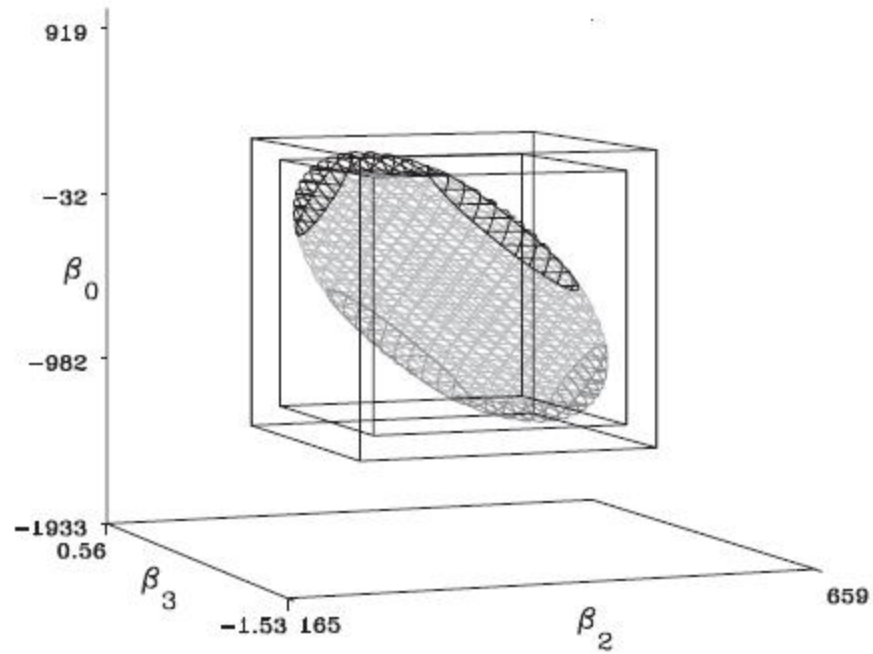
Uygun modelin bulunması için uygulanan bu özel yöntem pH ve $K'y_1$ içeren iki değişkenli bir modeli karşımıza çıkardı. Regresyon eşitliği

$$\hat{Y}_i = -507.0 + 412.0X_{i2} - 0.4871X_{i3} \quad (5.5)$$

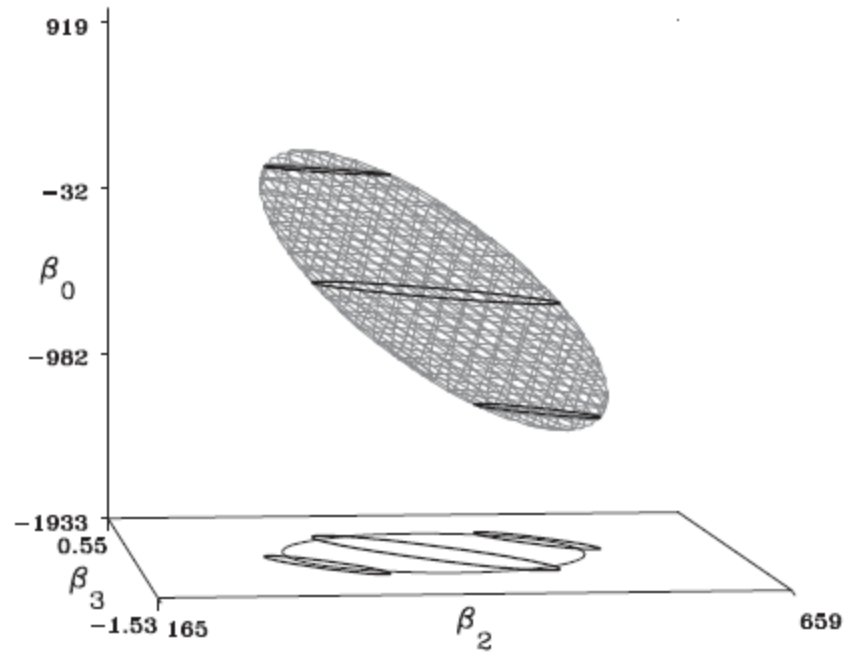
veya merkezileştirilmiş değişkenlerle belirtilmiş hali,

$$\hat{Y}_i = 1000.8 + 412.0(X_{i2} - 4.60) - .4871(X_{i3} - 797.62),$$

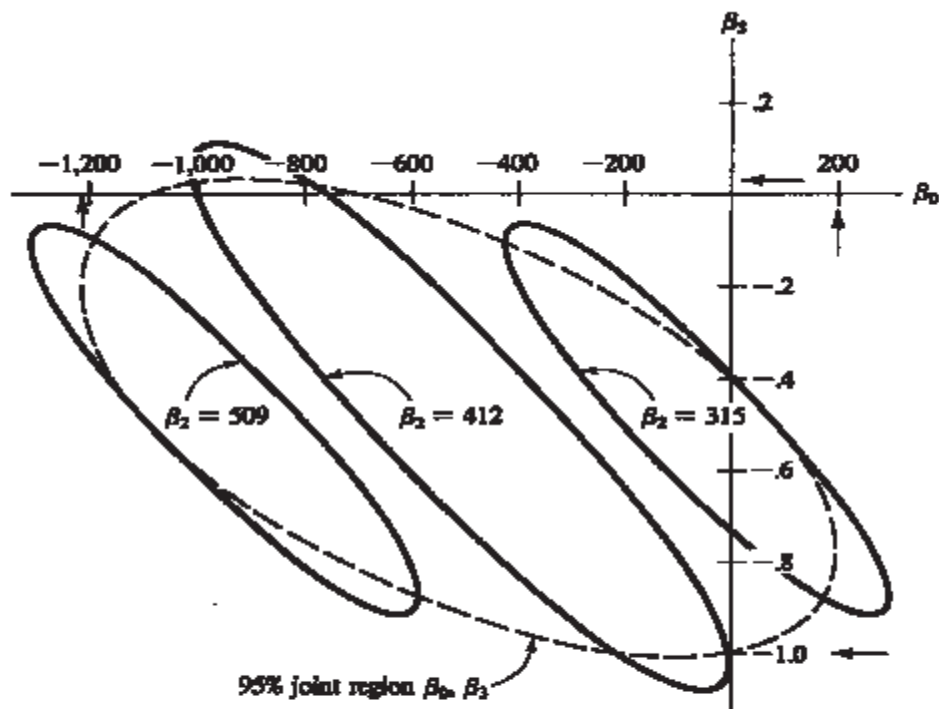
Burada $X_2 = pH$, $X_3 = K'y_1$ belirtmektedir. Bu eşitlik “havaya ilişkin” *BIOMASS*'i gözlemlenmiş değerlerindeki değişimin % 65'ni açıklamaktadır. Bu da $\hat{Y} = X\hat{\beta}$ dan hesaplanan kestirim değerleri *BIOMASS*'da ki değişimin % 65'ini açıkladığını göstermektedir ya da tersi bir deyişle artık kareler toplamı, $e'e$, genel *BIOMASS* modelinin düzeltilmiş kareler toplamının % 35'idir.



ŞEKİL 5.1. β_0 , β_2 ve β_3 için üç-boyutlu % 95 ortak güven bölgesi (elipsoid). Bonferroni güven aralıklarının kesişimi (iç kutu) ve Scheffe güven aralıklarının kesişimi (dış kutu).



ŞEKİL 5.2. Üç iki-boyutlu dilimlerin izdüşümünü gösteren, zeminde β_0 'ın üç değerine karşılık gelen; β_0 , β_2 ve β_3 için üç-boyutlu % 95 ortak güven bölgesi. Orta yerde dilimleri tanımlamak için seçilen β_0 'ın üç değeri ve β_0 için Bonferroni % 95 güven aralığının limitleri.



ŞEKİL 5.3. β_2 'nin üç değeri için iki-boyutlu ortak güven bölgesi ve β_2 ihmal edildiğinde β_0 ve β_3 için ortak güven bölgesi (noktalı çizgide gösterilmektedir.). Oklar burada β_0 ve β_3 için Benferroni güven aralıklarının kesişim sınırlarını göstermektedir.

GENEL DEĞERLENDİRME

- Linthurst araştırmanın asıl amacı bataklıklarda BIOMASS üretim miktarını etkileyen önemli toprak değişkenlerini belirlemektir.
- Bu amaç bize, isteğin nedensel ilişkiler sağladığını göstermiştir.
- Gözlemsel veriler nedensel ilişkileri kurmak için kullanılamaz.
- Örnek verilerinde; gözlemsel verilerle ilgili herhangi bir analiz, gözlemlenen ilişkilerin üzerine veya ilişki yapıları üzerine kurulmalıdır.
- Herhangi bir veri setinde korelasyonun var olması için birçok neden vardır.