

İki Yönlü Tablolarda Uyum Analizi Tekniğinin Kullanımı

Ensar BAŞPINAR¹

Mehmet MENDEŞ²

Geliş Tarihi : 14.02.2000

Özet: Bu çalışmada, iki yönlü tabloların analizinde yaygın olarak kullanılan χ^2 -testi, G-İstatistiği ve oranlara ait Z-Testine bir alternatif olarak Uyum Analizi (Correspondence Analysis) tekniği ele alınmıştır. Bu tekniğin uygulama aşamaları adım adım açıklanmış, diğer yöntemlere göre avantaj ve dezavantajları belirtilmiştir. Analiz tekniğinin adımlarına ait bir örnek, iki yönlü tablo üzerinde gösterilmiştir. Uyum analizi tekniği ile elde edilebilecek bilgi kalite ve miktarının diğer yöntemlerden daha yüksek olduğu sonucuna varılmıştır.

Anahtar Kelimeler: Uyum analizi, kategorik veri, iki yönlü tablo, χ^2 -testi, G-İstatistiği

The Usage of Correspondence Analysis Technique at the Contingency Tables

Abstract: In this study the Correspondence Analysis is handled as an alternative technique to χ^2 -test, G-Statistics and Z-test for proportions in analysis of contingency tables. The application stages of Correspondence analysis are explained step by step and advantages and disadvantages are determined with respect of the other techniques. The application stages were given in an example contingency table. As a result, the usage of this technique was more suitable than the other techniques given.

Key words: Correspondence analysis, categorical data, contingency table, χ^2 -test, G-statistics

Giriş

Üzerinde durulan özellik ya da özellikler bakımından, veriler farklı tiplerde elde edilebilir. Bazı araştırmalarda ölçmek, tartmak veya analiz etmek suretiyle elde edilirken bazı araştırmalarda saymak suretiyle ,var-yok şeklinde (binary), bazen de ya doğrudan kategorik bir şekilde elde edilir ya da elde edildikten sonra kategorize edilebilirler. Verilerin elde edilmiş şekillerinin bilinmesi, bunların değerlendirilmesi aşamasında kullanılacak istatistik tekniği ile doğrudan ilişkili olduğundan çok önemlidir. (Sokal ve Rohlf 1995)

Kategorik olarak elde edilmiş ya da elde edildikten sonra kategorize edilmiş verilerin analizi ile daha ziyade sosyal bilimlerde çok sık karşılaşılmaktadır. Bu tip veriler fen bilimlerinde başta tıp olmak üzere biyolojide, veteriner ve tarım bilimlerinde de oldukça yaygın bir şekilde rastlanılmaktadır. Bu şekilde elde edilmiş verilerin analiz edilmelerinde yaygın olarak kullanılan istatistik teknikler; χ^2 -istatistiği, Fisher'in Kesin Olasılık Hesaplanması, G-istatistiği ve Oran testi (Z-testi) dir. Ancak bu istatistik tekniklerinin kullanılabilmesi bazı şartlara bağlı olup, çoğu kez bu istatistiğin gerektirdiği şartlar sağlanamamakta yada bu şartlar sağlansa bile elde edilen bilgi çok genel kalmakta ve sonuçların yorumlanmasında bazı güçlüklerle karşılaşabilmektedir. Mesela bu tür veriler, iki yönlü tablo haline getirildikten sonra χ^2 -istatistiği ile analiz edilebilmesi, iki yönlü tablonun her bir hücresindeki beklenen frekanslarının en az 5 olması halinde güvenilir sonuçlar vermektedir. Böylece, üzerinde durulan özellik ya da özellikler bakımından elde edilen bilgi kaybının az olması sağlanabilmektedir. Diğer yandan,

oluşturulan iki yönlü tablonun bütün hücrelerinde beklenen frekansların 5 ve daha yukarısında olması durumunda bile elde edilen bilgi sadece satır değişkeni ile sütun değişkeninin birbirinden bağımsız olup olmadığına dair olup, oldukça yüzeysel bir bilgi vermektedir. Halbuki bir çok durumda araştırmacı, sadece satır ve sütun değişkeni arasında bir bağımlılığın olup olmadığını değil, aynı zamanda hem her bir değişkenin kendi seviyeleri (alt kategori) içindeki hem de değişkenlerin karşılıklı olarak seviyeleri arasındaki ilişkilerle ilgilenebilir. Bu durumda araştırmacının χ^2 -istatistiği ile amacına ulaşması oldukça zordur. Diğer yandan iki yönlü tablonun hücrelerindeki beklenen frekansların 5'in altına düşmesi durumunda söz konusu tablonun analizi için, iki yönlü tablonun çeşitli 2x2 tablolarına dönüştürülerek Fisher'in Kesin Olasılık Hesaplanması yaklaşımı kullanılabilir. Ancak hücrelerdeki beklenen frekanslardan bir kaçının sıfır olması durumunda bu testin uygulanması da araştırmacıya pek bir yarar sağlamayacaktır. Bu durumda G-istatistiğinin kullanılması düşünülebilir. Ancak söz konusu tablonun hücrelerinde bulunan sıfır sayısı kadar serbestlik derecesinden düşüleceği için elde edilecek bilginin güvenilirliği azalır. Hatta bazı durumlarda serbestlik derecesinin negatif olması da muhtemeldir. Böyle durumlar, bu testin kullanılabilirliğini sınırlandırmaktadır. Kullanılması mümkün olan bir diğer teknik olan oran testi ise olasılıkların (0, 1) açık aralığındaki durumlar için hesaplanabilir. Olasılığın tam olarak 1 veya 0 olması durumunda oran testi kullanılamaz. Sıfır frekanslı hücrelerin olması durumunda da oran testinin kullanılması ile elde edilecek sonuçlar yanıltıcı olurlar (Winer 1971, Snedecor ve Cochran 1980, Düzgüneş ve ark 1993).

¹ Ankara Üniv. Ziraat Fak. Zootečni Bölümü-Ankara

² Ankara Üniv. Çankırı Orman Fakültesi-Çankırı

Uyum analizi (CA) kategorik bir şekilde elde edilmiş yada elde edildikten sonra kategorize edilerek iki yönlü tablo haline getirilmiş verilerin hem satır ve sütun değişkenleri arasındaki ilişkilerin hem de her değişkenin kendi seviyeleri (alt kategorileri) arasındaki ilişkilerin açıklanmasına yardımcı olan ve aynı zamanda bu ilişkilerin düşük boyutlu bir uzayda grafik olarak da gösterilerek elde edilen sonuçların görsel olarak değerlendirilmesine imkan sağlayan bir tekniktir. Bu teknik, Çoklu uyum analizi (Optimal scaling), Temel Bileşenler Analizi ve Faktör Analizi ile ilişkili bir tekniktir. Bu tekniklerden Çoklu uyum analizi, Uyum Analizinin genel halidir. Uyum Analizi, üzerinde durulan özellik yada özellikler bakımından toplam varyasyonun parçalanmasında ve boyut indirgeme bakımından Temel Bileşenler Analizine benzerlik göstermektedir. Temel Bileşenler Analizinde X veri setindeki değişkenlerin, değişken sayısından daha az sayıda ve birbirinden bağımsız olan Temel Bileşenler elde edilebildiği, yani boyut indirgeme yapılabiliyorsa, benzer şekilde Uyum Analizinde de bir boyut indirgeme söz konusudur. Uyum Analizinde boyut sayısı, (satır sayısı-1) ya da (sütun sayısı-1)'den küçük olanı kadardır. Uyum Analizinde üzerinde durulan özellik ya da özellikler bakımından toplam varyasyonun ölçüsü olarak kullanılan toplam değişim (inertia) değeri de Temel Bileşenler Analizine benzer bir biçimde unsurlarına parçalanabilir. Uyum Analizinin Temel Bileşenler Analizinden en önemli farklılığı verilerin elde edilmiş şekillerinden kaynaklanmaktadır. Uyum Analizinde, veriler ya doğrudan kategorik bir şekilde elde edilmekte ya da elde edildikten sonra kategorize edilmektedir. Buna karşılık Temel Bileşenler Analizinde ise veriler, ölçmek, tartmak veya analiz etmek suretiyle elde edilmektedir (Nishisato 1980, Andersen 1990, Devilles ve Karcher 1991).

Uyum Analizi tekniği, özellikle bilgisayar yazılımlarının gelişmesiyle sosyal bilimlerde başta olmak üzere oldukça yaygın bir şekilde kullanılmaktadır. Bunun sebepleri ise kısaca aşağıdaki şekilde sıralanabilir:

- i. Bu tekniğin varsayımlardan bağımsız olması,
- ii. Diğer alternatiflerine göre kategorik verilerin analiz edilmesinde hem daha kolay hem de daha uygun olması,
- iii. Bu metotla aynı zamanda aynı uzayda satır ve sütun değişkenleri arasındaki çeşitli ilişkilerin grafik olarak gösterilebilmesi nedeniyle, sonuçların görsel olması ve kolayca yorumlanabilmesi,
- iv. Uyum Analizi, alternatifleri olabilecek tek değişkenli veya çok değişkenli yöntemlere nazaran uygulanmasının daha kolay olması, daha güvenilir bilgi elde edilebilmesi ve sonuçların daha geniş bir şekilde yorumlanmasına imkan sağlamasıdır (Deville ve Karcher 1991, Maulman 1998).

Bunlardan hareketle Uyum Analizinin, sosyal bilimlerde dışındaki birçok bilim alanında da çok rahatlıkla kullanılabileceği söylenebilir. Çünkü, birçok konuda elde edilen veriler, kategorik nitelikteki verilerdir. Özellikle biyolojik olaylarla ilgili çalışmaların birçoğunda durum

böyledir. Mesela belirli bir zararlı türünün ölüm oranının, farklı dozlardaki ilaçlara bağlı olup olmadığının araştırıldığı durumlarda, değişik ırklarından hayvan sayılarının buldukları bölgelere bağlı olup olmadığının belirlenmesinde, belirli bir ırkın dölerim özellikleri üzerine belirli bir hormonun değişik dozlarının etkili olup olmadığının belirlenmesinde, değişik göllerde yaşayan balık türlerinin dağılımlarının buldukları göllere bağlı olup olmadıklarının araştırılmasında, çeşitli buğday veya arpa varyetelerinin yaprakçık sayılarının dağılımları arasında fark olup olmadığının belirlenmesinde, çeşitli mikroorganizma türlerinin genomlarında bulunan adenin, guanin, cytosine ve timin amino asidi sayıları bakımından dağılımlarının araştırılması vb gibi çalışmalarda Uyum Analizi tekniği çok rahatlıkla kullanılabilir (Nishisato 1980, Greenacre 1984).

Bu çalışmada, yukarıda bahsedilen tekniklere göre daha detaylı bilgi verebilen, bu metotların yetersiz kaldıkları durumlarda çok rahatlıkla kullanılabilen ve herhangi bir ön şart gereksinim duymayan Uyum analizi(CA) tekniğinin özellikle biyolojik olaylarla ilgili çalışmalarda nasıl kullanılabileceği adım adım anlatılarak, bu metodun yukarıdaki alternatif tekniklerine göre üstünlükleri ele alınmış ve sonuçların nasıl yorumlanabileceği üzerinde durulmuştur.

Materyal ve Yöntem

Çalışmada, Uyum Analizi tekniğinin hesaplama adımlarını gerçek bir örnekle göstermek üzere, Öztürk ve Dellal (1999) tarafından "Border Leicester X Booroala (ff) X Merinos Melezi Koyunlarda Anestrus Dönemi Esnasında Farklı Dozlarda PMSG Uygulamasının Dölerim Üzerine Etkileri" ni araştırmak amacıyla yapmış oldukları çalışmada, PMSG dozu ile çeşitli doğum lipinde doğan kuzuların doğumdan itibaren bir aylık yaşa kadar ölenlerinin sayılarına ilişkin iki yönlü tablo materyal olarak kullanılmıştır.

Yöntemin hesaplama aşamaları "Minitab for Windows Version 12" İstatistik Paket Programı kullanılarak yapılmıştır.

Uyum Analizi, i seviyeli Z-değişkeni ile j seviyeli W-değişkenlerinin, kendi seviyeleri arasındaki ve içindeki ilişkileri açıklamaya yardımcı olan bir tekniktir. Z değişkeninin seviyeleri i adet satır ve W değişkeninin seviyeleri de j adet sütunda olmak üzere, her hangi bir Z x W tablosu i x j boyutlu bir iki yönlü tablo oluşturmaktadır. Bu tablonun her bir hücreindeki gözlenen frekansları f_{ij} ile, tablonun boyutunu $k=i \times j$ ile, satır toplamlarını $f_{i.}$, sütun toplamlarını $f_{.j}$ ve toplam gözlem sayısını da N ile gösterirsek; i x j boyutlu iki yönlü tablonun Uyum Analizi tekniği ile analizi (Z değişkeni 2 ve W değişkeninin de 3 seviyeli olduğu varsayılırsa) aşağıdaki işlem sıraları takip edilerek yapılır:

1. Tablonun her bir hücreindeki gözlenen frekansların marjinal gözlenme frekansları bulunur. (Satır ve sütun profilleri bulunur).

	W ₁	W ₂	W ₃	Σ
Z ₁	f ₁₁	f ₁₂	f ₁₃	f _{1.}
Z ₂	f ₂₁	f ₂₂	f ₂₃	f _{2.}
Σ	f _{.1}	f _{.2}	f _{.3}	f _{..}

Satır profilleri		
f ₁₁ /f _{1.}	f ₁₂ /f _{1.}	f ₁₃ /f _{1.}
f ₂₁ /f _{2.}	f ₂₂ /f _{2.}	f ₂₃ /f _{2.}

Sütun profilleri		
f ₁₁ /f _{.1}	f ₁₂ /f _{.2}	f ₁₃ /f _{.3}
f ₂₁ /f _{.1}	f ₂₂ /f _{.2}	f ₂₃ /f _{.3}

2. Bir satır ve sütunun seviye oranları (kategori oranları), söz konusu satırın, satır toplamının genel toplama ve söz konusu sütunun sütun toplamının da genel

toplama bölünmesi sonucu elde edilen ve satır veya sütunun ağırlık payı (mass) olarak adlandırılan satır ve sütun kategori oranları bulunur.

			Satır ağırlığı
f ₁₁	f ₁₂	f ₁₃	f _{1./N}
f ₂₁	f ₂₂	f ₂₃	f _{2./N}

			Sütun ağırlığı
f ₁₁	f ₁₂	f ₁₃	f _{.1/N}
f ₂₁	f ₂₂	f ₂₃	f _{.2/N}
			f _{.3/N}

3. Kategorilerin birbirlerine göre uzaklıkları hesaplanır. Bu işlem, farklı yaklaşımlarla yapılabilmektedir. Mesela; χ^2 -uzaklığı veya öklit uzaklıkları gibi. Bu çalışmada χ^2 -uzaklıkları esas

alınmıştır. Bu uzaklıklar; $\chi^2_{dij} = \frac{(f_{ij} - f'_{ij})^2}{f_{ij}}$ şeklinde

hesaplanmaktadır. Bu ifadede; f_{ij} : her bir hücredeki gözlenen frekansları, f'_{ij} : her bir hücredeki beklenen frekansları göstermektedir. $f'_{ij} = f_{i.} \cdot f_{.j} / N$ şeklinde hesaplanır.

4. χ^2 -uzaklıklarından yararlanılarak tablonun toplam χ^2 -değeri; $\chi^2 = \sum \chi^2_{ij}$ şeklinde hesaplanır.

Hesaplanan χ^2 değeri, toplam gözlem sayısına bölünerek tablonun genel değişim (inertia) değeri bulunur.

6. Elde edilen genel değişim değerinden yararlanılarak, tablodaki her bir değişkene ilişkin kategorilerinin birbirlerinden olan uzaklıklarının daha az boyutla (eksenle) açıklanıp açıklanamayacağı hususunda bir fikir edinilebilir. Toplam değişim değeri (aynı zamanda öz-değerler olarak da bilinmektedir) her bir boyut için, satır profilleri ile sütun profilleri arasındaki kanonik korelasyon olarak adlandırılan tekil değer (singular value) karesidir. Toplam değişim değeri, aslında tablodaki varyansın bir ölçüsüdür. Bu değerden yararlanılarak ele alınan tablodaki varyasyonun kaç boyutla açıklanabileceği belirlenir. Bunun için toplam değişim değerinin her bir boyuttaki (eksendeki) payını yani, her bir boyutun açıklayabildiği varyasyon payının bulunması gerekir. Bunun için, önce tekil değerlerin (singular values) bulunması gerekmektedir. Tekil değerlerin bulunması aşağıdaki adımlar izlenerek yapılabilmektedir.

6.1 Oluşturulan iki yönlü tablo $i \times j$ boyutlu bir matrise aktarılır (F_{ij}),

$$F = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \end{bmatrix}$$

6.2. Satır ve Sütun toplamlarının yer aldığı r ve c vektörleri bulunur. Bunlar;

$r = Fv$ ve $c = uF$ (u, v , elemanları 1 olan vektörler olup, $u; r \times 1$, $v; c \times 1$ boyutludur) şeklinde elde edilebilir.

$$r = \begin{bmatrix} f_{1.} \\ f_{2.} \end{bmatrix} \quad c = \begin{bmatrix} f_{.1} \\ f_{.2} \\ f_{.3} \end{bmatrix}$$

6.3 r ve c vektörleri $D_r = \text{diag}(r)$ ve $D_c = \text{diag}(c)$ şeklinde köşegen matrise dönüştürülürler.

$$D_r = \begin{bmatrix} f_{1.} & 0 \\ 0 & f_{2.} \end{bmatrix} \quad D_c = \begin{bmatrix} f_{.1} & 0 & 0 \\ 0 & f_{.2} & 0 \\ 0 & 0 & f_{.3} \end{bmatrix}$$

6.4. F matrisinin Uyum analizi sonuçlarının daha az sayıdaki boyutla gösterimini elde etmek için, gerekli olan her bir satırın ve her bir sütunun koordinat eksenindeki koordinatlarının belirlenmesini sağlayacak X ve Y matrislerinin bulunması gerekir. Bunun için;

6.4.1. F 'nin X matrisi şeklindeki gösteriminin elde edilmesi gerekir. Bunu elde etmek için, F , D_r ve D_c matrislerinden yararlanılarak; $E = D_r^{-1/2} F D_c^{-1/2}$ elde edilir. Bu E matrisi, toplam değişim değerinin her bir boyuttaki açıklanabilen varyasyon payını gösterecek olan boyutsal değişim değerlerinin elde edilmesinde kullanılır. Bunun için de E matrisinin, tekil değere ayrıştırma (Singular Values Decomposition=svd) yöntemi kullanılarak (K, Λ, V)=svd(E) şeklinde ayrıştırılması gerekir. Bu işlem aşağıdaki adımlar izlenerek yapılır.

6.4.1.1. $A = E'E$ kare matrisi bulunur.

6.4.1.2. A matrisinin öz-değerleri bulunur ($|A - \lambda I| = 0$ 'dan yararlanılarak).

6.4.1.3. Her bir öz-değere karşılık gelen öz-vektörler bulunur,

6.4.1.4. Öz-vektörler yardımıyla V matrisi, öz-değerlerin kareköklerinin köşegen matrise dönüştürülmesi ile de Λ matrisi elde edilir,

6.4.1.5. K matrisinin her bir sütununun oluşturduğu k_i vektörleri;

$$k_i = \frac{1}{\sqrt{\lambda_i}} E v_i \text{ şeklinde elde edilir } (v_i, i. \text{ öz-vektördür})$$

6.4.2. Satır profillerinin apsilerini ve Sütun profillerinin ordinatlarını gösteren X ve Y matrisleri;

$X = N^{1/2} D_r^{-1/2} K \Lambda$ $Y = N^{1/2} D_c^{-1/2} V \Lambda$ şeklinde kolayca hesaplanır.

Yukarıdaki K ve V matrisleri $K'K=V'V=I$ şartını sağlamalıdır.

7. X ve Y matrislerinin elemanlarından yararlanılarak, satır ve sütun değişkenleri arasındaki ilişkiler grafiksel olarak gösterilebilir. Bu grafikte, her değişkenin her seviyesi koordinat sisteminde ayrı bir nokta olarak temsil edilir. Bu noktalardan bir birine yakın olanların, üzerinde durulan özellikler bakımından birbirlerine benzer oldukları, birbirinden uzak olanların da farklı oldukları anlaşılır. Aynı zamanda, çizilen grafikte satır değişkeninin seviyelerinin X-ekseni boyunca sıralanmaları yani X-ekseni üzerinde yada ona çok yakın olmaları, satır değişkeninin seviyelerinin üzerinde durulan özellik bakımından birbirlerinden bağımsız oldukları, dolayısıyla birbirlerini etkilemedikleri anlamına gelmektedir. Benzer durum, sütun değişkenleri için de Y-ekseni baz alınarak geçerlidir.

Bulgular ve Tartışma

Çalışmada kullanılan iki yönlü tablo, Çizelge 1'de verilmiştir.

Çizelge 1. PMSG uygulanan ve uygulanmayan koyunlarda doğum-1. ay arasında ölen kuzu sayıları ¹⁾

PMSG dozu (IU)	W faktörü				Toplam
	Tekiz	İkiz	Üçüz	Dördüz	
0	1	1	0	0	2
250	3	6	0	0	9
300	7	10	12	4	33
400	4	3	11	0	18
600	0	4	6	2	12
Toplam	15	24	29	6	74

1) Öztürk ve Derral 1999'dan alınmıştır.

Çizelge 2. Satır profilleri ve satır ağırlık oranları

Doz (IU)	Satır profilleri				Satır ağırlık oranı
	Tekiz	İkiz	Üçüz	Dördüz	
0	1/2= .500	1/2= .500	0/2= .000	0/2= .000	2/74= .027
250	3/9= .333	6/9= .667	0/9= .000	0/9= .000	9/74= .122
300	7/33= .212	10/33= .303	12/33= .364	4/33= .121	33/74= .446
400	4/18= .222	3/18= .167	11/18= .611	0/18= .000	18/74= .243
600	0/12= .000	4/12= .333	6/12= .500	2/12= .167	12/74= .162

Çizelge 3. Sütun profilleri ve sütun ağırlık oranları

Doz (IU)	Sütun profilleri			
	Tekiz	İkiz	Üçüz	Dördüz
0	1/15=0.067	1/24=0.042	0/29=0.000	0/6=0.000
250	3/15=0.200	6/24=0.250	0/29=0.000	0/6=0.000
300	7/15=0.467	10/24=0.417	12/29=0.414	4/6=0.667
400	4/15=0.267	3/24=0.125	11/29=0.379	0/6=0.000
600	0/15=0.000	4/24=0.167	6/29=0.207	2/6=0.333
Sütun ağırlık oranı	15/74=0.203	24/74=0.324	29/74=0.392	6/74=0.081

PMSG'nin değişik dozlarının (Z faktörü) doğumdan itibaren bir aylık yaşa kadar ölen kuzuların doğum tipine (W faktörü) etkisinin Uyum Analizi (Correspondance Analysis) tekniği ile araştırılması amacıyla, W faktörünün 4 (Tekiz, İkiz, Üçüz ve Dördüz) seviyesi ile Z faktörünün 5 (0 IU, 250 IU, 300 IU, 400 IU ve 600 IU) seviyesinin ele alındığı böyle bir çalışmada, eğer χ^2 -testi, G-İstatistiği ve Fisher'in Kesin Olasılık Hesaplanması test tekniklerinden yararlanılırsa, W faktörünün seviyelerini Z faktörünün seviyelerinden bağımsız olup olmadıklarına karar verilebilir. Fakat bu karar araştırmacının esas araştırmak istediği konuya bir cevap teşkil etmez. Çünkü araştırmacı W faktörünün hangi seviyesinin Z faktörünün hangi seviyesi üzerine etki ettiğini merak etmektedir. Bu amacını gerçekleştirmek üzere, W faktörünün her bir seviyesinde Z faktörünün seviyelerine ait oranları karşılaştırarak veya Z faktörünün her bir seviyesinde W faktörünün seviyelerine ait oranları ikiye ikiye karşılaştırarak ulaşmak isteyebilir. Ancak bu durumda da, ya gözlemlerin aynı popülasyonlardan alınmış rastgele örneklerden elde edilmiş olduğunu, ya da söz konusu gözlemlerin farklı popülasyonlardan alınmış rastgele örnekler olduklarını varsaymak durumundadır. Bu varsayımların her ikisi de her zaman ele alınan örneklerin yapısına uygun olmayabilmektedir. Bu nedenle varılan hükümlerde bir miktar gerçek durumdan uzaklaşma olabilmektedir. Bir başka sakıncalı durum da, gözlenen oranların 1 veya 0 çıkması halinde ortaya çıkmaktadır. Çünkü oranın 1 veya 0 çıkması oran testinin kullanılmasını kısıtlar. Bu sakıncaları önlemek üzere, herhangi bir varsayım gerektirmeyen ve araştırmacının amacına da uygun olan, Uyum Analizi tekniği rahatlıkla kullanılabilir. Bu teknik araştırmacıya daha fazla bilgi verebileceği gibi yorumlamalarını da daha kolay yapmasına imkan vermektedir.

Çizelge 1'deki verilerin Uyum Analizi tekniği ile irdelenme aşamaları aşağıdaki gibidir.

i) İki yönlü tablonun satır, sütun profilleri ve ağırlık oranları bulunur. Bu değerler sırasıyla Çizelge 2 ve 3'de verilmiştir.

ii) χ^2 -uzaklıkları $\chi^2_{d_{ij}} = \frac{(f_{ij} - f'_{ij})^2}{f'_{ij}}$ ifadesine göre hesaplanır

(Çizelge 4).

Çizelge 4. χ^2 -Uzaklıkları

Doz (IU)	Tekiz	İkiz	Üçüz	Dördüz	Toplam
0	0.872	0.190	0.784	0.162	2.008
250	0.758	3.252	3.527	0.730	8.267
300	0.014	0.046	0.067	0.655	0.783
400	0.034	1.380	2.207	1.459	5.080
600	2.432	0.003	0.358	1.084	3.877
Toplam	4.110	4.871	6.943	4.091	20.016

Toplam χ^2 değeri, (5-1)x(4-1)=12 serbestlik dereceli χ^2 tablo değeri ile karşılaştırılırsa, Doğum-1ay arasında ölen kuzuların doğum tiplerinin koyunlara uygulanan PMSG dozlarından bağımsız olduğu kararı verilebilir. Ancak, χ^2 -testi ile varılan bu karar pek güvenilir değildir. Çünkü, χ^2 -testinin yapılabilmesi, ancak her bir hücredeki beklenen frekansların 5 ve daha büyük olması halinde güvenilirdir. Halbuki bu örnekte 15 hücrede beklenen frekans 5'ten az, 6 hücrede de 1'den daha az bulunmuştur. Kaldı ki bu sonuç güvenilir olsa bile, araştırmacının asıl merak ettiği soruya yani, "Kuzularda ölüm oranının PMSG dozlarına göre değişip değişmediğini, eğer değişiyorsa hangi dozda en az olmaktadır?" sorularına cevap vermekten uzaktır. Benzer durum G-İstatistiği için de geçerlidir. Eğer, kuzuların ölüm oranları PMSG dozu seviyelerine göre ikişer ikişer karşılaştırılmak istense (iki oranın karşılaştırılmasına ilişkin Z-Testi kullanılarak), bu örnekte söz konusu testi kullanmak da pek mümkün değildir. Çünkü oranların bir kısmı 0'dır. Öztürk ve Dellal (1999) bu hükümden hareketle, sadece satır ağırlık oranlarına göre PMSG dozlarını karşılaştırmakla yetinmiş, doz ve doğum tiplerine göre ölüm oranları hakkında bilgi edinemedişlerdir. Halbuki Uyum Analizi ile durumun gerçekten böyle olup olmadığı daha doğru bir yaklaşımla ortaya konulabilir.

iii) W x Z iki yönlü tablosu F matrisine aktarılır ve bundan yararlanarak, Satır ve sütun toplamalarını içeren D_r ve D_c matrisleri oluşturulur.

$$F = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 3 & 6 & 0 & 0 \\ 7 & 10 & 12 & 4 \\ 4 & 3 & 11 & 0 \\ 0 & 4 & 6 & 2 \end{bmatrix} \quad D_r = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 & 0 \\ 0 & 0 & 33 & 0 & 0 \\ 0 & 0 & 0 & 18 & 0 \\ 0 & 0 & 0 & 0 & 12 \end{bmatrix}$$

$$D_c = \begin{bmatrix} 15 & 0 & 0 & 0 \\ 0 & 24 & 0 & 0 \\ 0 & 0 & 29 & 0 \\ 0 & 0 & 0 & 6 \end{bmatrix}$$

iv) E matrisinin bulunması,

$$E = D_r^{-1/2} F D_c^{-1/2} = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 & 0 \\ 0 & 0 & 33 & 0 & 0 \\ 0 & 0 & 0 & 18 & 0 \\ 0 & 0 & 0 & 0 & 12 \end{bmatrix}^{-1/2} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 3 & 6 & 0 & 0 \\ 7 & 10 & 12 & 4 \\ 4 & 3 & 11 & 0 \\ 0 & 4 & 6 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 15 & 0 & 0 & 0 \\ 0 & 24 & 0 & 0 \\ 0 & 0 & 29 & 0 \\ 0 & 0 & 0 & 6 \end{bmatrix}^{-1/2} = \begin{bmatrix} .182574 & .144338 & .000000 & .000000 \\ .258199 & .408248 & .000000 & .000000 \\ .314627 & .355335 & .387905 & .284268 \\ .243432 & .144338 & .481457 & .000000 \\ .000000 & .235702 & .321634 & .235702 \end{bmatrix}$$

v) $A = E' E$ matrisinin bulunması,

$$A = E' E = \begin{bmatrix} .182574 & .258199 & .314627 & .243432 & .000000 \\ .144338 & .408248 & .355335 & .144338 & .235702 \\ .000000 & .000000 & .387905 & .481457 & .321634 \\ .000000 & .000000 & .284268 & .000000 & .235702 \end{bmatrix}$$

$$= \begin{bmatrix} .182574 & .144338 & .000000 & .000000 \\ .258199 & .408248 & .000000 & .000000 \\ .314627 & .355335 & .387905 & .284268 \\ .243432 & .144338 & .481457 & .000000 \\ .000000 & .235702 & .321634 & .235702 \end{bmatrix}$$

$$= \begin{bmatrix} .258249 & .278696 & .239247 & .089438 \\ .278696 & .390152 & .283138 & .156566 \\ .239247 & .283138 & .485719 & .186079 \\ .089438 & .156566 & .186079 & .136364 \end{bmatrix}$$

vi) Tekil değerlere ayrıştırma (singular values decomposition) metodu ile A matrisinden hareketle, K, Λ ve V matrislerinin bulunması,

vi.1. A matrisinin öz-değerlerinin bulunması,

$$|A - \lambda I| = 0 = \begin{vmatrix} .258249 - \lambda & .278696 & .239247 & .089438 \\ .278696 & .390152 - \lambda & .283138 & .156566 \\ .239247 & .283138 & .485719 - \lambda & .186079 \\ .089438 & .156566 & .186079 & .136364 - \lambda \end{vmatrix} = 0$$

determinantından elde edilecek 4.dereceden denklemin çözülmesi ile, $\lambda_1 = 1.0000000$, $\lambda_2 = .1781152$, $\lambda_3 = .0767465$ ve $\lambda_4 = .0156219$ öz-değerleri elde edilir.

vi.2. Öz-değerlerin karekökleri alınıp köşegen matrisle dönüştürülerek Λ matrisi elde edilir,

$$\Lambda = \begin{bmatrix} \sqrt{1.0000000} & 0 & 0 & 0 \\ 0 & \sqrt{.1781152} & 0 & 0 \\ 0 & 0 & \sqrt{.0767465} & 0 \\ 0 & 0 & 0 & \sqrt{.0156219} \end{bmatrix}$$

$$= \begin{bmatrix} 1.0000000 & 0 & 0 & 0 \\ 0 & .4220370 & 0 & 0 \\ 0 & 0 & .2770315 & 0 \\ 0 & 0 & 0 & .1249874 \end{bmatrix}$$

vi.3. Öz-değerlere karşılık gelen öz-vektörler;

$$\lambda_1=1.0000000 \text{ için } v_1 = \begin{bmatrix} -.450225 \\ -.569495 \\ -.626013 \\ -.284747 \end{bmatrix}$$

$$\lambda_2=.1781152 \text{ için } v_2 = \begin{bmatrix} -.406725 \\ -.549569 \\ .702350 \\ .198122 \end{bmatrix}$$

$$\lambda_3=.0767465 \text{ için } v_3 = \begin{bmatrix} -.514982 \\ .318345 \\ -.261201 \\ .751814 \end{bmatrix} \text{ ve}$$

$$\lambda_4=.0156219 \text{ için } v_4 = \begin{bmatrix} -.605529 \\ .521829 \\ .215839 \\ -.560752 \end{bmatrix} \text{ olarak bulunduğundan}$$

sonra V matrisi;

$$V = \begin{bmatrix} -.450225 & -.406725 & -.514982 & -.605529 \\ -.569495 & -.549569 & .318345 & .521829 \\ -.626013 & .702350 & -.261201 & .215839 \\ -.284747 & .198122 & .751814 & -.560752 \end{bmatrix}$$

şeklinde elde edilir.

vi.4. K matrisini elde etmek üzere k_i ($i=1,2,3,4$) vektörleri;

$$k_i = \frac{1}{\sqrt{\lambda_i}} E v_i \text{ ifadesi yardımıyla hesaplanır,}$$

$$k_1 = \frac{1}{\sqrt{1.0000000}} \begin{bmatrix} .182574 & .144338 & .000000 & .000000 \\ .258199 & .408248 & .000000 & .000000 \\ .314627 & .355335 & .387905 & .284268 \\ .243432 & .144338 & .481457 & .000000 \\ .000000 & .235702 & .321634 & .235702 \end{bmatrix}$$

$$\begin{bmatrix} -.450225 \\ -.569495 \\ -.626013 \\ -.284747 \end{bmatrix} = \begin{bmatrix} -.164399 \\ -.348743 \\ -.667792 \\ -.493197 \\ -.402694 \end{bmatrix}$$

$$k_2 = \frac{1}{\sqrt{.1781152}} \begin{bmatrix} .182574 & .144338 & .000000 & .000000 \\ .258199 & .408248 & .000000 & .000000 \\ .314627 & .355335 & .387905 & .284268 \\ .243432 & .144338 & .481457 & .000000 \\ .000000 & .235702 & .321634 & .235702 \end{bmatrix}$$

$$\begin{bmatrix} -.406725 \\ -.549569 \\ .702350 \\ .198122 \end{bmatrix} = \begin{bmatrix} -.236408 \\ -.507011 \\ .008493 \\ .246008 \\ .220217 \end{bmatrix}$$

$$k_3 = \frac{1}{\sqrt{.0767465}}$$

$$\begin{bmatrix} .182574 & .144338 & .000000 & .000000 \\ .258199 & .408248 & .000000 & .000000 \\ .314627 & .355335 & .387905 & .284268 \\ .243432 & .144338 & .481457 & .000000 \\ .000000 & .235702 & .321634 & .235702 \end{bmatrix}$$

$$\begin{bmatrix} -.514982 \\ .318345 \\ -.261201 \\ .751814 \end{bmatrix} = \begin{bmatrix} -.091335 \\ -.005707 \\ .120621 \\ -.389809 \\ .319620 \end{bmatrix}$$

$$k_4 = \frac{1}{\sqrt{.0156219}}$$

$$\begin{bmatrix} .182574 & .144338 & .000000 & .000000 \\ .258199 & .408248 & .000000 & .000000 \\ .314627 & .355335 & .387905 & .284268 \\ .243432 & .144338 & .481457 & .000000 \\ .000000 & .235702 & .321634 & .235702 \end{bmatrix}$$

$$\begin{bmatrix} -.605529 \\ .521829 \\ .215839 \\ -.560752 \end{bmatrix} = \begin{bmatrix} -.099663 \\ .160349 \\ -.228464 \\ .090037 \\ .170413 \end{bmatrix}$$

şeklinde bulunan k_i vektörlerinden yararlanılarak K matrisi oluşturulur,

$$K = \begin{bmatrix} -.164399 & -.236408 & -.091335 & -.099663 \\ -.348743 & -.507011 & -.005707 & .160349 \\ -.667792 & .008493 & .120621 & -.228464 \\ -.493197 & .246008 & -.389809 & .090037 \\ -.402694 & .220217 & .319620 & .170413 \end{bmatrix}$$

$K'K = V'V = I$ olduğuna dikkat edilmelidir.

vi.5. Satır ve sütun değişkenlerinin koordinatlarının belirlenmesi,

$$X = N^{1/2} D_r^{-1/2} K \Lambda$$

$$X = 74^{1/2} \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 & 0 \\ 0 & 0 & 33 & 0 & 0 \\ 0 & 0 & 0 & 18 & 0 \\ 0 & 0 & 0 & 0 & 12 \end{bmatrix}^{-1/2}$$

$$\begin{bmatrix} -.164399 & -.236408 & -.091335 & -.099663 \\ -.348743 & -.507011 & -.005707 & .160349 \\ -.667792 & .008493 & .120621 & -.228464 \\ -.493197 & .246008 & -.389809 & .090037 \\ -.402694 & .220217 & .319620 & .170413 \end{bmatrix}$$

$$\begin{bmatrix} 1.0000000 & 0 & 0 & 0 \\ 0 & .4220370 & 0 & 0 \\ 0 & 0 & .2770315 & 0 \\ 0 & 0 & 0 & .1249874 \end{bmatrix}$$

$$= \begin{bmatrix} -1 & -.60690 & -.15391 & -.07577 \\ -1 & -.61357 & -.00453 & .05747 \\ -1 & .00537 & .05004 & -.04276 \\ -1 & .21051 & -.21896 & .02282 \\ -1 & .23080 & .21988 & .05289 \end{bmatrix}$$

$$\text{ve } Y = N^{1/2} D_c^{-1/2} V \Lambda$$

$$Y = 74^{1/2} \begin{bmatrix} 15 & 0 & 0 & 0 \\ 0 & 24 & 0 & 0 \\ 0 & 0 & 29 & 0 \\ 0 & 0 & 0 & 6 \end{bmatrix}^{-1/2}$$

$$\begin{bmatrix} -.450225 & -.406725 & -.514982 & -.605529 \\ -.569495 & -.549569 & .318345 & .521829 \\ -.626013 & .702350 & -.261201 & .215839 \\ -.284747 & .198122 & .751814 & -.560752 \end{bmatrix}$$

$$\begin{bmatrix} 1.0000000 & 0 & 0 & 0 \\ 0 & .4220370 & 0 & 0 \\ 0 & 0 & .2770315 & 0 \\ 0 & 0 & 0 & .1249874 \end{bmatrix}$$

$$= \begin{bmatrix} -1 & -.38126 & -.31688 & -.16810 \\ -1 & -.40727 & .15486 & .11453 \\ -1 & .47350 & -.11559 & .04309 \\ -1 & .29365 & .73144 & -.24614 \end{bmatrix}$$

olarak bulunurlar.

Gerek X matrisinin gerekse Y matrisinin ilk sütunu daima 1'lerden oluşur. Bunun sebebi $\lambda_1=1$ olacak şekilde ayrıştırma işleminin yapılmasıdır. Dolayısıyla, X ve Y matrislerinin ilk sütunları değerlendirmeye alınmazlar.

vi.6. Yukarıda bulunan sonuçlar yardımıyla Uyum analizi tekniğine esas teşkil eden iki yönlü tablonun, analiz tablosu Çizelge 5'deki gibi oluşturulur. Bu analiz tablosuna dikkat edilirse, χ^2 istatistiğinin önemsiz olduğu görülebilir. Dolayısıyla satır ve sütun değişkenlerinin birbirlerinden bağımsız olduğu söylenir. Bundan dolayı satır ve sütun değişkenlerinin etkilerine pratikte genellikle ayrı ayrı bakılmaktadır. Ancak Uyum Analizi tekniğinde hem sadece satır değişkeni hem de sadece sütun değişkeni kendi seviyeleri arasındaki farklılıklar bakımından ele alındıkları gibi, ikisi birlikte de değerlendirilebilmektedir.

Çizelge 5. Uyum analizi tablosu

Boyutlar	Tekil değer	Boyutların toplam değişim değerini açıklama oranları			χ^2 -Değeri (12 S.D.'li)	Önemlilik düzeyi	Kanonik korelasyonlar
		Boyutlardaki	Açıklama payı	Ekleme payı			
1	.4220370	.1781152	.6585066	.6585066	20.016	.067	-.086
2	.2770315	.0767465	.2837382	.9422448			
3	.1249874	.0156219	.0577555	1.0000000			
Toplam		.2704835					

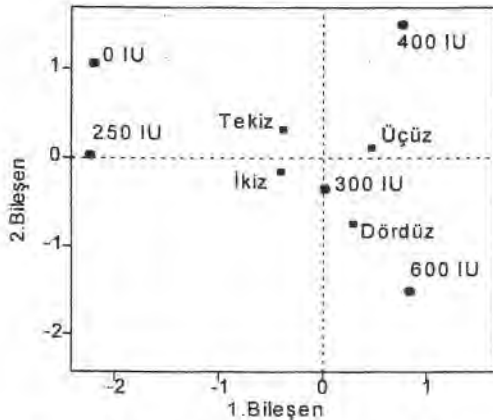
Değişkenlerin birbirlerinden bağımsız ($P>0.05$ olması halinde) olmalarına rağmen hücrelerdeki gözlem sayılarının düşüklüğü böyle bir incelemeyi gerekli kılmaktadır. Çünkü toplam gözlem sayısının artırılması halinde değişkenlerin birbirine bağımlı çıkması

muhtemeldir. Bu durumda araştırmacı iki değişkenin interaksyonu (satır ve sütun değişkenlerinin karşılıklı etkileşimi) hakkında da bilgi sahibi olmak ister. Satır ve sütun değişkenleri ilk önce ayrı ayrı değerlendirilecek olursa;

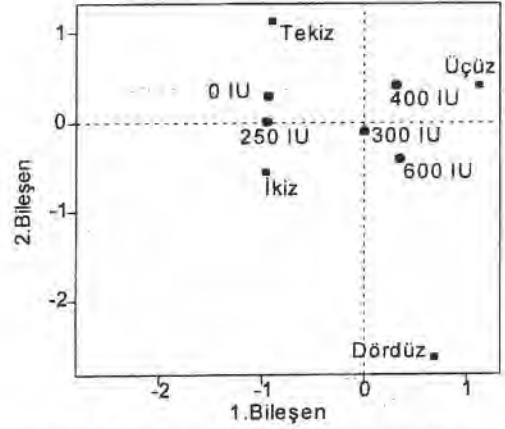
Sadece sütun değişkeni yani doğum tipi göz önüne alınarak bir karşılaştırma yapıldığında (Şekil 1);

1. 300 IU PMSG dozunun en çok üçüzlerde ölüm frekansını artırdığını,
2. Dördüzlerde ölüm frekanslarının yine 300 IU PMSG uygulamasında en yüksek olduğu,
3. İkizlerde 300 IU PMSG uygulamasının yaklaşık dördüzler kadar olduğu (çünkü, ikizlerin ve dördüzlerin 300 IU dozuna olan uzaklıkları hemen hemen aynıdır.),
4. 300 IU dozunun tekizlerde de ölüm frekansını etkilediğini söylemekle birlikte, diğer doğum tiplerine nazaran bu etkinin daha az olduğunu,
5. 250 IU dozunun hemen hiçbir doğum tipinde ölüm frekanslarına önemli bir etkisinin olmadığını,
6. 400 IU dozunun 300 IU dozundan daha az olmakla birlikte üçüzlerde ölüm frekanslarını etkilediğini,
7. 600 IU dozunun ise, en çok dördüzlerde ölüm frekansının artırdığını, buna karşılık en az tekizler üzerinde etkili olduğunu,
8. 0 IU dozunun sadece tekiz ve ikizlerde çok az bir etkiye sahip olduğunu (doğum tipi ve PMSG dozlarının birbirlerine göre χ^2 -uzaklıklarına bakılarak),
9. Ayrıca, satır değişkeninin (PMSG dozlarının) her bir seviyesi ile sütun değişkeninin (doğum tipi) her bir seviyesinin Şekil 1'deki buldukları bölgeler (++, +, -, -) göz önüne alınarak da doz ve doğum tiplerine göre, ölüm frekansları arasındaki ilişkiler incelenebilir. Mesela, 400 IU dozu ile üçüz doğum tipi Şekil 1'de (++) yani 1. bölgede yer aldıklarından bunlar arasında pozitif yönde bir ilişkinin olduğu söylenebilir.

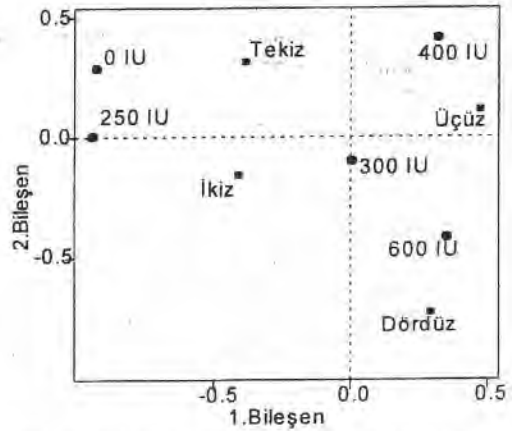
Bütün bu değerlendirmelerde, sadece sütun değişkeninin belirli bir seviyesi göz önüne alınıp satır değişkeninin çeşitli seviyeleri birbirleriyle karşılaştırılmaktadır.



Şekil 1. Sütunlara göre uyum analizi sonuçları



Şekil 2. Satırlara göre uyum analizi sonuçları



Şekil 3. Satır ve sütunlara göre uyum analizi sonuçları

Satır değişkeni (Şekil 2) esas alınarak bir değerlendirme yapılırsa;

1. 400 IU dozunun en çok üçüzlerde ölüm frekanslarını etkilediği,
2. 300 IU dozunun üçüzler üzerinde 400 IU dozundan daha az bir ölüme sebep olduğu,
3. 300 IU dozunun ikizlerde de üçüzlerdekine benzer bir etkiye sahip olduğunu,
4. Tekizlerin en fazla 0 IU dozunda ölüm frekanslarının fazla olduğunu, 250 IU, 300 IU ve 400 IU dozlarının ise, yaklaşık olarak aynı etkiye sahip olduklarını,
5. 600 IU dozunun kendisine en yakın olan ikiz ve üçüzlerde yaklaşık olarak aynı etkiye sahip olduğu,
6. Dördüzlerin ölüm frekansları üzerine ise, hemen hiçbir dozun etkili olmadığını söylemekle birlikte, kendisine en yakın olan 600 IU dozu ile ilişkili olduğu söylenebilir.

Bu yorumlarda da sadece satır değişkeninin belirli bir seviyesi göz önüne alınarak sütun değişkeninin çeşitli seviyeleri birbirleriyle karşılaştırılırken satır değişkeninin diğer halleri göz ardı edilmektedir.

Satır ve sütun değişkenleri birlikte (Şekil 3) ele alındığında;

1. 400 IU dozunun üçüzlerde ölüm frekansını etkilediğini ve (++) bölgesinde bulunmalarından dolayı pozitif yönde ilişkili oldukları söylenebilir.

2. 600 IU dozunun daha ziyade dördüzlerde ölüm frekansı üzerinde etkili olduğu ve (+-) bölgesinde bulunmaları sebebiyle dozun artması durumunda dördüzlerde ölüm frekanslarında bir azalmanın olabileceğini, benzer bir etkiye daha az olmakla birlikte 300 IU dozunun da sebep olduğu iler sürülebilir.

3. İkizlerde ölüm frekansları üzerine hemen hiçbir dozun etkili olmadığı söylenebilir.

4. Tekizlerde ölüm frekanslarının en çok 0 IU dozundan etkilendiği söylenebilir.

Sonuç

Sonuç olarak, yukarıda yapılan üç farklı yorumlamadan en sağlıklı bilgiyi veren ve araştırmacının amaçlarına en uygun olanı, satır ve sütun değişkenlerinin karşılıklı etkileşimlerinin göz önüne alındığı en son yorumlamalardır. Bu durum, Uyum Analizi tekniğinin alternatiflerine göre üstünlüklerini ortaya koymaktadır. Çünkü χ^2 -testinde, Fisher'in Kesin Olasılık Hesaplanmasında ve oranlara ait Z-testinde bu yorumlamalar yapılamamaktadır.

Araştırmacı isterse, sadece satır değişkenini göz önüne alarak (Şekil 1), sütun değişkeninin seviyelerini veya sadece sütun değişkenini göz önüne alarak (Şekil 2) satır değişkeninin seviyelerini karşılaştırabilir. Fakat bunun

istatistik açıdan her hangi bir sakıncası olmamakla birlikte, bilgi kaybına sebep olacağı da göz ardı edilmemelidir.

Kaynaklar

- Andersen, E.B. 1990. The Statistical Analysis of Categorical Data. Springer-Verlag, New York. USA.
- Deville, J. and W.Karcher, 1991. Applied Multivariate Analysis. Sar and Environmental Studies. Kluwer Acad.Pub., Dordrecht, The Netherlands.
- Düzgüneş, O., T.Kesici, F.Gürbüz, 1993. İstatistik Metodları. Ank.Üniv.Zir.Fak.Yayınları: 1291. Ders Kitabı: 369. II.Baskı.
- Greenacre, M.J. 1984. Theory and Applications of Correspondence Analysis. London Academic Press. UK.
- Meulman, J.J. 1998. Optimal Scaling Methods for Graphical Multivariate Data Analysis. XIII Symposium on Computational Statistics, Bristol.
- Nishisato, S. 1980. Analysis of Categorical Data: Dual Scaling and Its Applications. Toronto, Univ. Of Toronto Press. Canada.
- Öztürk, A.K. ve G.Dellal, 1999. Border Leicester x Booroola (ff) x Merinos Melezi Koyunlarda Anestrus Dönemi Esnasında Farklı Dozlarda PMSG Uygulamasının Döl Verimi Üzerine Etkileri. A.Ü.Z.F. Tarım Bilimleri Dergisi, 5(3), 35-39.
- Snedecor, G.W. and W.G. Cochran, 1980. Statistical Methods. Seventh Ed. The Iowa State University Press. Ames, Iowa, U.S.A. 507 s.
- Sokal, R.R. and F.J.Rohlf, 1995. Biometry. The Principles and Practice of Statistics in Biological Research. Third Ed. W.H. Freeman and Co., New York. 887 s.
- Winer, B.J. 1971. Statistical Principles in Experimental Design. Second Ed. McGraw-Hill Book Co., New York. 907 s.