SERİ B CİLT 52 SAYI 1 2002

İSTANBUL ÜNİVERSİTESİ

ORMAN FAKÜLTESİ

D E R G İ S İ

F.1

DOĞA BİLİMLERİNDE Kİ-KARE BAĞIMSIZLIK TESTİ VE TEK YÖNLÜ VARYANS ANALİZİ TESTİ UYGULANMASINDA SPSS PROGRAMI KULLANIMI

Y.Doç.Dr. Arif KUBAT[1]
Y.Doç.Dr. Tülay AYAŞLIGİL[2]

Kısa Özet

Doğa bilimlerinde elde edilen verilerin, istatistiksel olarak değerlendirmesinde kullanılan analizlerden; nitel veriler için çok sık olarak kullanılan 'Ki-Kare Analizi' ile nicel değişkenler için kullanılan 'Tek Yönlü Varyans Analizi' teknikleri açıklanmıştır. Başlangıçta bu uygulama analizlerinin teorisi hakkında kısaca bilgilendirme yapılmıştır. Günümüzde çevre ve ormancılık alanında yapılacak araştırmalarda yardımcı olması amacıyla, bu çalışmanın örnekleri doğa bilimlerinden seçilmiştir. Bitki türlerinin yetişme ortamı koşullarından yükseklik ve ıslaklık faktörlerine bağlı olarak yayılışlarına ait verilerin bilgisayar ortamında SPSS paket programı kullanılarak analizi ile bilgisayar çıktılarının tablo yorumlamaları ve anlamlılık değerlendirmesi ayrıntılı olarak yapılmıştır. Bu tablolardaki analizlerin kolay anlaşılabilmesi için grafiksel ifade teknikleri kullanılmıştır.

Anahtar Kelimeler: Ki-Kare Bağımsızlık Testi, Tek Yönlü Varyans Analizi, SPSS

THE APPLICATION OF INDEPENDENT Kİ-SQUARE TEST AND ONE-WAY ANALYSIS OF VARIANCE IN THE NATURAL SCIENCES USING SPSS PROGRAM

Abstract

The analysis methods used for the statistical evaluation of the data obtained in the natural sciences, "Ki-Square Analysis'', which is used very often for qualitative data and "One-Way Analysis of Variance" used for the quantitative data have been explained. At first,information on the theory of such application analysis methods was provided briefly. In order to help research to be made on environment and foresty today, the examples of this study have been chosen from natural sciences. The data on the distrubution of woody plants

1) İÜ Sosyal Bilimler Meslek Yüksek Okulu
2) YTÜ Mimarlık Fakültesi Şehir ve Bölge Planlama Bölümü

Yayın Komisyonuna Sunulduğu Tarih: 09.08.2002

as based on the elevation and moisture, the two of the growing environment conditions, were analyzed using SPSS packet program in the computer and comments on tables and significance values have been performed in details. Graphical comment techniques have been used in order to facilitate the comprehension of the analysis given in the tables.

Keywords: Independent Ki-Square Test, One-Way Analysis of Variance, SPSS

1. GİRİŞ

Örnekleme yöntemiyle toplanan veriler düzenlenerek çeşitli istatistiksel metodlarla analiz edilir. Veri analizinin amacı, eldeki problemi çözmeye yardımcı olacak bilgiyi sağlamaktır. Bu aşamalar araştırma işleminin önemli bir bölümünü oluştururlar. Çünkü, iyi bir istatistik analiz yapılmaksızın isabetli karar vermek mümkün değildir. Uygulamada çoğu zaman örneklemeden elde edinilen bilgiler yardımıyla, ana kütle parametreleri hakkında bir karara varmaya çalışılmaktır.

Araştırmalarda örneklem sayısı büyüdükçe, yapılan analizlerin güvenirliği artar. Çok sayıdaki verileri analiz etmede bilgisayarın önemi daha da artırmıştır. Verileri değerlendirmek için istatistik analiz testleri yapan istatistik paket programlar vardır, bunların en yaygın kullanılanı SPSS paket programıdır. İstatistik metodlar genellikle her alandaki bilimsel araştırmada verileri analiz etmekte kullanılmaktadır. Bu makalede doğa bilimi verilerini analiz etmek için bilgisayar paket programındaki bazı istatistik test teknikleri açıklanmaktadır.

2. Kİ-KARE BAĞIMSIZLIK TESTİ

2.1 Ki-Kare Bağımsızlık Testi Analizi

Ki-Kare bağımsızlık testi analizi iki değişken arasında bir ilişki bulunup bulunmadığını veya gözlenen ve beklenen frekans dağılımları arasındaki farkların rasgele hatalardan olup olmadığını kontrol etmede kullanılmaktadır. Böyle bir test örneğinde gözlenen bir frekans dağılımının Binom, Poison veya normal dağılıma uygun olup olmadığına, şayet frekanslar iki özelliğe göre dağılım gösteriyorsa, bu iki özelliğin birbirinden bağımsız olup olmadığına karar vermede yardımcı olur.

Genellikle karşılaşılan araştırma sorularından birincisi, iki değişkenin birbiri ile ilişkili olup olmadığıdır. Örneğin bir araştırmacı orman yangınına kişisel nedenlerden sebep verenlerin eğitim düzeyi ile ilişkisinin var olup, olmadığını belirlemek isteyebilir. İki değişken arasında ilişki yoksa yani birinin dağılımı hiçbir şekilde diğerinin dağılımına bağımlı değilse, bu iki değişken birbirinden bağımsızdır denir. İki değişken arasında ilişki yoksa; değişkenlerden birinin özelliğinin değerini bilmek, diğer değişkenin özelliğinin değerinin bilinmesine imkan vermez.

Örneğin ağacın yaşı ile boy uzunluğu birbirinden bağımsız ise, belirli bir ağacın boy uzunluğunun bilinmesi, o ağacın yaşını tahmin etmede bize bilgi veremez.

Veriler genel olarak aşağıdaki gibi bir basamak tablosunda gösterilir. Her bir değişkenin tablodaki bir değerini temsil eden n_{ij} gözlem sayısı, i inci satır ve j inci sütunun kesiştiği yerdeki hücreye yazılır. Hücre içerisindeki değer, gözlenen hücre frekansı olarak adlandırılır ve genellikle

f_{ij} sembolü ile gösterilir, yani $f_{ij} = n_{ij}$ dir. Gözlenen f_{ij} hücre frekansı, 1. nci değişkenin i inci basamağı ile 2. nci değişkenin j inci basamağındaki veri sayısını gösterir (Tablo 1).

Tablo 1 : Ki –Kare Gözlem Değerleri Kontenjans Tablosu

		2. inci Değişken						
		Basamaklar						**Satır toplam**
1. inci Değişken	**Basamaklar**	**1**	**2**	**...**	**j**	**....**	**c**	
	1	n_{11}	n_{12}	....	n_{1j}	....	n_{1c}	n_{1+}
	2	n_{21}	n_{22}	....	n_{2j}	....	n_{2c}	n_{2+}
	⋮	⋮	⋮	....	⋮	⋮	⋮	⋮
	i	n_{i1}	n_{i2}	....	n_{ij}	....	n_{ic}	n_{i+}
	⋮	⋮	⋮	....	⋮	⋮	⋮	⋮
	r	n_{r1}	n_{r2}	....	n_{rj}	....	n_{rc}	n_{r+}
	Sütun toplam	n_{+1}	n_{+2}	....	n_{+j}	....	n_{+c}	n

Hipotezler:

H_0 = İki değişkenin basamakları birbirinden bağımsızdır.

H_a = İki değişkenin basamakları birbirinden bağımsız değildir.

2.2 Test İstatistiğinin Hesaplanması

H_0 hipotezinin doğru olduğu varsayımı kabul edilir. Beklenen frekansları hesaplamada şu olasılık kanunu kullanılır. İki olay birbirinden bağımsız ise, birlikte vuku bulma olasılığı, iki olayın ayrı ayrı vuku bulma olasılıklarının çarpımına eşittir.

İki olay birbirinden bağımsız ise, i inci satır ve j inci sütundaki beklenen olayın olasılığı, bu olayların olasılıkları birbiriyle çarpıldığında n birimli bir örnek için bulunur.

Böylece ij hücresi olasılığı : P (ij hücresi olayı) = $(n_i / n) \cdot (n_j / n)$

Bu ij hücresindeki beklenen frekans υ_{ij} değerini elde etmek için toplam örnek hacmi olan n ile çarpılır : $\upsilon_{ij} = n\,(n_i / n) \cdot (n_j / n)$ buradan $\upsilon_{ij} = (n_i \cdot n_j / n)$ elde edilir.

Buradaki bağımsızlık bir olayın meydana geliş frekansı ile diğer bir olayın meydana geliş frekansından etkilenmemesi anlamındadır. Bunu iki olayın bağımsızlığını olasılık formülü bağıntısından şöyle açıklayabiliriz:

$P(A \cdot B) = P(A) \cdot P(B)$

Bu ifadenin sağlanması yani rasgele olayın birlikte meydana gelme olasılığının, bu olayların ayrı ayrı meydana gelme olasılıklarının çarpımına eşit olması gerekmektedir. Bu bağımsızlığın karşıtı ise,

$P(A \cdot B) = P(A) \cdot P(B / A)$ olması gerekir.

Ki-Kare testi uygulama araştırmalarında genellikle karşılaşılan hatalardan biri de küçük teorik frekanslarla çalışmadır. Kesikli dağılım olan Binom'un normal dağılıma yaklaşması nedeni aynı şekilde, Ki-kare dağılımı kullanılmasıyla da açıklanmaktadır. Ancak teorik frekansların 10' dan büyük olması şartı ile sınırlandırılmıştır. Uygulamada bu şartın biraz daha hafifletilerek, serbestlik derecesi 5' den büyük olduğu durumlarda 10' dan büyük olmanın 5' den büyük olma şekline indirildiği görülmektedir. Böylece serbestlik derecesi 5' den az veya 10' dan küçük bir teorik frekans olmamalı, eğer varsa uygun gruplama imkanı araştırılmalıdır. Gruplama yapılamıyorsa hipergeometrik dağılımına dayanan ve Fisher'in kesin Ki-kare testi uygulanır.

x ve y değişkenlerinin birlikte meydana gelme olasılığı bileşik olasılıktır. i inci satır ve j inci sütunun bulunduğu hücrede frekansın olma olasılığı p_{ij} olarak tarif edilir. Yukarıdaki kontenjans tablosunda yer alan n_{ij} frekans değerleri toplam frekansı gösteren n değerine bölündüğünde p_{ij} olasılıkları elde edilir (Tablo 2).

Tablo 2 : Ki –Kare Olasılık Değerleri Kontenjans Tablosu

		Y- değişkeni basamakları						**Satır toplam olasılıkları**
		1	2	...	j	...	c	
X-değişkeni basamakları	1	p_{11}	p_{12}	...	p_{1j}	...	p_{1c}	p_{1+}
	2	p_{21}	p_{22}	...	p_{2j}	...	p_{2c}	p_{2+}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	i	p_{i1}	p_{i2}	...	p_{ij}	...	p_{ic}	p_{i+}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	r	p_{r1}	p_{r2}	...	p_{rj}	...	p_{rc}	p_{1+}
Sütun toplam olasılıkları		p_{+1}	p_{+2}	...	p_{+j}	...	p_{+c}	1

p_{ij} değerlerinin her biri, bu iki değişkenin birlikte olma olasılığını göstermektedir. Bu bileşik olasılıkların satır ve sütun toplamları alınarak marjinal değerler elde edilir. Satır ve sütun marjinal değerlerin toplamı 1'e eşittir (AGNESTI, 1996).

Sıfır ilişki olması, değişkenlerin bileşik olasılık dağılımına bağlıdır. Değişkenlerin bileşik dağılım basamaklarındaki durum sayısına bağlı olarak tahmin edilebilirse bu değişkenlerin birbirinden bağımsız olduğu söylenir. Bir kontenjans tablosundaki herhangi bir hücrenin beklenen frekans olasılığı bu hücrenin bulunduğu satır ve sütun toplamlarının, genel toplama bölünmesiyle elde edilir. Hücrelerin gözlenen frekansları, bileşik olasılık kanununa göre hesaplanan frekanslara eşit olduğunda Ki-kare değeri sıfırdır (GARSON, 1999).

Gözlenen hücrelerin olasılıkları p_{ij} ile gösterilirse bütün hücreler için $p_{ij} = p_{i+} \cdot p_{+j}$ olması durumunda, iki değişkenin birbirinden bağımsız olduğunu iddia eden sıfır hipotezi kabul edilir (BISHOP, 1991). Bağımsızlık durumunda hangi değişkenin bağımsız değişken hangi değişkenin bağımlı değişken olduğu önemli değildir. Yani iki değişkende birbirinden bağımsızken, değişken nitelikleri karşılıklı olarak yer değiştirebilir, simetriden bahsedilebilinir.

Ki-kare bağımsızlık testinde, 1'e yakın serbestlik dereceleri için mutlaka Yates düzeltmesi kullanılması, 10'dan küçük teorik frekansın bulunmaması gerektiği daima hatırlanmalıdır. Serbestlik derecesi 5 civarında ise teorik frekanslar 10 ile 5 arasında olabilir ama asla 5'den küçük olmamalıdır, var ise birbirine yakın şıklar gruplanmalıdır (YOĞURTÇUGİL, 1978).

2.3 Uygulamalı Örnek Açıklaması

Çok nadir görülen özel bir bitki türünün, yetişme ortamı, bölge ve yüksekliğe göre sayısal miktarları aşağıdaki tabloda verilmiştir. Bu bitkinin yetişmesinde, bölge ile yükseklik arasında anlamlı bir ilişkinin var olup olmadığı araştırılır. Bu hipotezden hareketle aşağıdaki tabloyu kullanarak Toroslar ve Karadeniz örneğini incelemeye çalışalım (Tablo 3, 4).

Tablo 3 : Elde Edilen Verilerin Ki-Kare Kontenjans Tablosu

	Yükseklik			**Satır toplam**
	1000 m. ve altı	**1001-2000 m.**	**2001 m. ve üstü**	
Toroslar	30	8	2	40
Karadeniz	10	48	2	60
Sütun toplam	40	56	4	100

Görüldüğü gibi tablo değerleri arasında 5'den küçük değerler mevcuttur. Bu tabloyu yeniden gruplamaya gitmemiz gerekecektir. 1001-2000 m. ile 2001 m. ve üstü sütunlarını birleştirip tek sütun olarak dikkate alınır, yani yeni sütun 1001 m. ve üstü olarak ifade edilir (Tablo 4).

Tablo 4 : Düzeltilmiş Ki-Kare Kontenjans Tablosu (Tablo 3'ün Düzeltilmişi)

	Yükseklik (metre)		**Satır toplam**
	1000 m. ve altı	**1001 m. ve üstü**	
Toroslar	30 (16)	10 (24)	40
Karadeniz	10 (24)	50 (36)	60
Sütun toplam	40	60	100

Tablodaki parantez içindeki değerler beklenen değerleri göstermektedir. Bu sayılar bulunduğu satır ve sütun toplamlarının çarpımının genel toplama bölünmesiyle elde edilmiştir.

$$\upsilon_{ij} = \frac{n_i \cdot n_j}{n}$$

Ki-kare test değeri ise aşağıdaki formül kullanarak hesaplanmıştır.

$$\chi^2_h = \sum_{i=1}^{n_r} \sum_{J=1}^{n_c} \frac{(f_{ij} - \upsilon_{ij})^2}{\upsilon_{ij}}$$

Gözlenen değerler (f_{ij})	Beklenen değerler (υ_{ij})	(f_{ij}- υ_{ij})	(f_{ij}- υ_{ij})^2	$\frac{(f_{ij}-\upsilon_{ij})^2}{\upsilon_{ij}}$
30	16	14	196	12,25
10	24	14	196	8,16
10	24	14	196	8,16
50	36	14	196	5,44

$$\chi^2_{hesap} = 34,01$$

İstatistik kitaplarında Ki-kare tablo değeri ise serbestlik derecesi ve anlamlılık düzeyine göre belirtilmiştir. Ki-kare tablosu serbestlik derecesinin hesabı ise, satır ve sütun sayılarından 1 çıkarılarak, birbirleriyle çarpımı sonucunda serbestlik derecesi elde edilir.

Bunu şöyle ifade edebiliriz:

$v = (r-1) . (c-1)$ Buradaki örneğimize uygularsak : $v = (2-1) . (2-1) = 1$ olur.

Ki-kare tablo değeri $\chi^2_1 (0,05) = 3,841$

$$\chi^2_{tablo} < \chi^2_{hesap}$$

Ki-kare hesaplanan değeri, tablo değerinden büyük olduğundan (3,841<34,01) yetişme ortamı, bölge ile yükseklik arasında anlamlı ilişki vardır denir. Bulduğumuz sonucu yorumladığımızda, bu bitki türü Toroslarda çoğunlukla 1000 m. ve altında, Karadeniz'de ise çoğunlukla 1001 m. ve üstünde yetişmektedir diyebiliriz.

2.4 Ki-Kare Testinin Bilgisayar Uygulaması

Bir örnek alan içerisinde bulunan toplam 7591 adet, iki farklı grup bitki türünün, yetişme ortamının yükselti durumuna göre yayılışı araştırılmak istenmiştir. Burada kullanılan yükseklik kriterleri şöyledir : 1-)1000 m. ve altı, 2-)1001-2000 m., 3-) 2001-3000 m. Bu iki farklı grup bitki türünün yayılışında bitkilerin yüksekliğe bağlı olup olmadığı anlamlılık ilişkisi test edilmiştir (Tablo 5, Şekil 1).

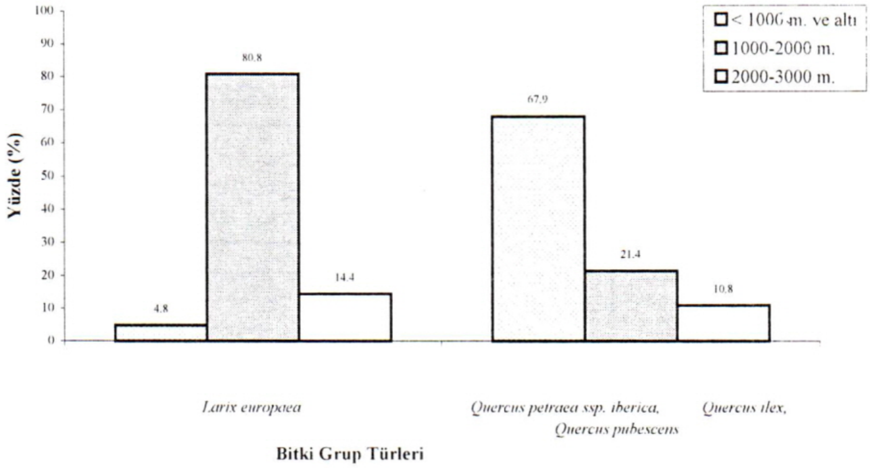
Tablo 5 : Yükseltiye Göre Bitki Grup Türlerinin, Bilgisayar Çıktısı Ki-Kare Tablosu

			YÜKSELTİ			Satır toplam
			<1000 m. ve altı	1000-2000m.	2000-3000 m.	
TÜR	*Larix europaea*	Sayı(n)	192	3258	580	4030
		Satır(%)	4,8%	80,8%	14,4%	100,0%
		Sütun(%)	5,1%	74,5%	50,8%	43,6%
		Toplam(%)	2,1%	35,2%	6,3%	43,6%
	Quercus petraea ssp. *iberica* *Quercus ilex* *Quercus pubescens*	Sayı(n)	3540	1115	562	5217
		Satır(%)	67,9%	21,4%	10,8%	100,0%
		Sütun(%)	94,9%	25,5%	49,2%	56,4%
		Toplam(%)	38,3%	12,1%	6,1%	56,4%
Sütun toplam		Sayı(n)	3732	4373	1142	9247
		Satır(%)	40,4%	47,3%	12,3%	100,0%
		Sütun(%)	100,0%	100,0%	100,0%	100,0%
		Toplam(%)	40,4%	47,3%	12,3%	100,0%

Ki-kare analiz örneğinin bilgisayar çıktısında anlamlılık düzeyi P= 0,001 bulunmuştur. Buna göre P=0,0001***< 0,001 olduğundan bitki türü ile yükseklik arasında ileri derecede anlamlı ilişki vardır. *Larix europaea* çoğunlukla 1000-2000 m. arasında yetişmekte (% 80,8); fakat *Quercus petraea* ssp. *iberica*, *Quercus ilex*, *Q. pubescens* türleri ise çoğunlukla 1000m. ve altında yetişmektedir (% 67,9). Bu iki grup bitki türünün 2000-3000m. arasındaki yetişme dağılımları düşük düzeyde olup birbirine yaklaşıktır. Yüzdeleri sırasıyla %10,8 ve % 14,4 dür.

Ki-Kare tablosunun satır yüzde dağılım değerlerini kullanarak grafik çizersek, incelemeye çalıştığımız farklılığı daha açık ve kolay göstermiş oluruz. Grafikte görüldüğü gibi *Larix europaea* 1000-2000 m. de % 80,8 ile en yüksek değere sahip ve diğer bitki grubunda ise 1000 m. ve altında % 67,9 ile en yüksek değeri görülmektedir (Şekil 1).

Buradaki bilgisayar çıktısından elde edilen sonucu yorumladığımızda; *Larix europaea* bitki türünün oluşturduğu grup çoğunlukla (% 80,8) 1001-2000 m. de, diğer grup bitki türüde çoğunlukla (% 67,9) 1000 m. ve altı yükseklikte yayıldığı görülmektedir. 2001-3000 m. de ise iki grup bitki türüde düşük oranda ve hemen hemen aynı olduğu görülmüştür, yüzde oranları sırasıyla % 14,4 ve % 10,8 dir.



Şekil 1 : Yükselti durumuna göre bitki türlerinin yayılışı

3. TEK YÖNLÜ VARYANS ANALİZİ TESTİ (One-Way)

3.1 Tek Yönlü Varyans Analizi

Varyans analizinde tek yönlü uygulamada, bir değişkenin basamaklarına göre ana kütle ortalamalar arasında anlamlı bir fark olup olmadığını test edilir. Basamak sayısı ikiden çok olduğu zaman başvurulan bir test tekniğidir, basamak sayısı iki olduğunda zaten t-testi kullanılır. Fakat testin güvenirliliği açısından söz konusu değişkenin basamaklarının çok sayıda olmamasına dikkat edilmesi gerekmektedir. Örneğin, 5 ayrı bölgedeki aynı ağaç türlerinin ortalama boyları arasında farkı testinde, burada her bir bölge farklı basamak kabul edilerek analiz yapılır. Burada test edilecek hipotez kurulurken k-sayıda ana kütle ortalamasının birbirine eşit olduğu $\mathbf{H_0}$ hipotezi, eşit olmadığı ise $\mathbf{H_a}$ hipotezidir. Sembolik olarak aşağıdaki gibi ifade edilir:

$$\mathbf{H_0}: \mu_1 = \mu_2 = \mu_3 = \ldots\ldots\ldots\ldots\ldots\ldots \mu_k$$

$$\mathbf{H_a}: \mu_1 \neq \mu_2 \neq \mu_3 \neq \ldots\ldots\ldots\ldots\ldots\ldots \mu_k$$

Analiz için her ana kütleden $n_1, n_2, n_3 \ldots\ldots n_k$ sayıda örnekleme tesadüfi olarak seçilir. Bu örneklemlerin ortalamaları olan $\bar{x}_1, \bar{x}_2, \bar{x}_3 \ldots\ldots \bar{x}_k$ hesaplanarak her ana kütlenin varyansının tahmini yapılır.Buradan k sayıda örnek varyanslarının ortalaması S_o^2 hesaplanır.

Tablo 6'da 5 basamaklı (k=5) bir model gösterilmiştir. 1. 2. 3. 4. 5. olarak adlandırdığımız bu kategorilerin her birinde tesadüfi olarak seçilmiş 8 gözlem bulunduğundan $n_1 = n_2 = n_3 = n_4 = n_5 = 8$ olur. Tabloda x_{ij} ile i inci satırdaki j'inci sütundaki birimin değerini ifade etmekte ve T_{1+}, T_{2+}, T_{3+}, T_{4+}, T_{5+} ile sütunların toplamını, T_{++} ile de genel toplamı belirtmektedir (Tablo 6).

Tablo 6: Değişkenin Kategori Değerleri

	Değişkenin Kategorileri						**Genel ortalama**
	1.	**2.**	**3.**	**4.**	**5.**	**Genel toplam**	
	x_{11} x_{12} x_{13} x_{14} x_{15} x_{16} x_{17} x_{18}	x_{21} x_{22} x_{23} x_{24} x_{25} x_{26} x_{27} x_{28}	x_{31} x_{32} x_{33} x_{34} x_{35} x_{36} x_{37} x_{38}	x_{41} x_{42} x_{43} x_{44} x_{45} x_{46} x_{47} x_{48}	x_{51} x_{52} x_{53} x_{54} x_{55} x_{56} x_{57} x_{58}		
Sütun toplamları	T_{1+}	T_{2+}	T_{3+}	T_{4+}	T_{5+}	T_{++}	
Aritmetik ortalama	$\overline{x}_1$	$\overline{x}_2$	$\overline{x}_3$	$\overline{x}_4$	$\overline{x}_5$		$\overline{x}$

Tablo- 6 daki değerlerin hesaplanması aşağıda gösterilmiştir.

$$T_{1+} = x_{11} + x_{12} + x_{13} + x_{14} + x_{15} + x_{16} + x_{17} + x_{18}$$

$$T_{2+} = x_{21} + x_{22} + x_{23} + x_{24} + x_{25} + x_{26} + x_{27} + x_{28}$$

$$T_{3+} = x_{31} + x_{32} + x_{33} + x_{34} + x_{35} + x_{36} + x_{37} + x_{38}$$

$$T_{4+} = x_{41} + x_{42} + x_{43} + x_{44} + x_{45} + x_{46} + x_{47} + x_{48}$$

$$T_{5+} = x_{51} + x_{52} + x_{53} + x_{54} + x_{55} + x_{56} + x_{57} + x_{58}$$

$$T_{++} = T_{1+} + T_{2+} + T_{3+} + T_{4+} + T_{5+} + T_{6+} + T_{7+} + T_{8+}$$

$$\overline{x}_1 = \frac{T_{1+}}{8}$$

Basamaklarının varyanslarının ortalaması olan ortalama varyans s_o^2 nin hesabı şu şekildedir.

$$s_o^2 = \frac{\sum_{j=1}^{n_1} (x_{1j} - \overline{x}_1)^2 + \sum_{j=1}^{n_2} (x_{2j} - \overline{x}_2)^2 + \sum_{j=1}^{n_3} (x_{3j} - \overline{x}_3)^2 + \dots\dots\dots + \sum_{j=1}^{n_8} (x_{8j} - \overline{x}_8)^2}{n_1 + n_2 + n_3 + n_4 + n_5 + n_6 + n_7 + n_8 - 8}$$

Basamak ortalamalarının varyansını eğer denek sayıları eşitse n ile çarparak S_m^2 elde edilebilir.

$$S_m^2 = n \frac{\sum_{i=1}^{k} (\overline{X_i} - \overline{X_{......}})^2}{k-1}$$

Basamaklarının denek sayıları birbirinden farklı olduğunda ise aşağıdaki formül yardımıyla hesaplanır.

$$S_m^2 = \frac{\sum_{i=1}^{k} n_i (\overline{X_i} - \overline{X_{......}})^2}{k-1}$$

Bu hesaplamaların neticesinde S_o^2 ve S_m^2 değerleri, ortalamalar arasında anlamlı fark olup olmadığını test etmekte kullanılır. S_m^2 değeri S_o^2 değerine bölünmesiyle F - test değeri elde edilir. Daha önceden de belirtildiği gibi, ana kütle ortalamaları arasında fark olduğu durumda S_m^2 değeri S_o^2 dan daima büyüktür. S_o^2 değerine "örnekler-içi ortalama varyans''olarak tanımlanır ve rasgele nedenlerin ortaya çıkardığı değişkenliği ölçer. S_m^2 değerine de "örnekler arası ortalama varyans" denir ve ortalama farklarından ileri gelen değişkenliği ölçer.

Böylece F-test istatistik değeri:

$$F_{hesap} = \frac{S_m^2}{S_o^2}$$ şeklinde ifade edilir.

Bulunan F_{hesap} değeri , F-dağılım tablosundaki k-1 ve $\Sigma k_i - 1$ serbestlik dereceleri ile belirlenen F_{tablo} değerinden büyük olduğunda H_o hipotezi red edilecektir. Yani, bu basamak ortalamaları arasında anlamlı fark vardır denir, H_a hipotezi kabul edilmiş olur.

F-test istatistik değeri hesaplamasının daha pratik bir yöntemi olan Tablo 6 daki toplam değerleri kullanarak da hesaplanır. Bu hesaplamaların tümünü bir tablo, şeklinde göstermek mümkündür. İşte bu tabloya "Tek Yönlü Varyans Analizi Tablosu" denir ve aşağıdaki şekilde ifade edilir (Tablo 7).

Tablo 7 : Tek Yönlü Varyans Analizi Tablosu

	Kareler toplamı	**Serbestlik derecesi**	**Varyans**	**F-değeri**
Örnekler arası	$\sum \frac{(T_{i+})^2}{n_i} - \frac{(T_{++})^2}{N}$	k-1	S_m^2	$F_{hesap} = \frac{S_m^2}{S_o^2}$
Örnekler içi	$\sum\sum(X_{ij})^2 - \sum \frac{(T_{i+})^2}{n_i}$	N-k	S_o^2	
Genel toplam	$\sum\sum(X_{ij})^2 - \sum \frac{(T_{++})^2}{N}$	N-1		

Hata terimleri ile ilgili varsayımlarımıza göre ana kütleler arasındaki farklılık yalnız ortalamaların farklılığından ileri gelmektedir. r tane ana kütlenin ortalamaları arasındaki farklılığı bulmak için ikişer ikişer karşılaştırarak incelemek de mümkündür.

Bu amaçla r'nin 2'şerli kombinasyonu kadar t-testi yapılması gerekir. Bu testlerin hepsinde farkların önemli olmadığı sonucuna varılırsa faktör kategorileri arasında fark olmadığı anlaşılır. Fakat testin bu şekilde yapılmasında önemli sakıncalar vardır. Bu testlerin hepsinde H_o'nin tüm çiftleri için $\mu_i = \mu_j$ hipotezlerinin kabul edilir olma olasılığı, eğer $\alpha = 0,05$ alınırsa, $(0.95)^{10}=0,60$ böylece on tane t-testi birlikte kabul edilebilir olmasının anlamlılık düzeyi $1\text{-}0,60=0,40$ olarak hesaplanır. Bu anlamlılık düzeyi ise 0,05 değerinden oldukça büyüktür. Sonuçta bu test sonucunun güvenirliliği azdır diyebiliriz. İşte varyans analizi yöntemi bu olumsuzluğu önlemiş olur.

Varyans analizi sonucunda grup ortalamaları arasında anlamlı farklar bulunduğunda yani H_o hipotezinin red edildiğinde o zaman hangi gruplar arasındaki farklılık önemli olduğu araştırılabilinir. İşte çok sayıda gruptan oluşan varyans analizinde, analiz sonucu anlamlı çıktığında, bu anlamlılığın hangi gruplar arasından oluştuğunu bulmak için Scheffe testi yöntemi ile bulunarak kolayca belirtilebilinir (ERSOY / ERBAŞ 1992).

3.2 Tek Yönlü Varyans Analizi Bilgisayar Uygulaması

Tek yönlü varyans analizinin bilgisayar uygulamasında bir örnek alan içerisinde toplam 9034 adet ve sekiz farklı bitki türünden oluşan denekler, toprağın ıslaklık durumuna ait beşli ölçek kriterlerine göre incelenmiştir (POLUNIN/WALTERS 1985). Kullanılan ölçeğin ıslaklık kriterleri şöyledir: 1-Çok kuru, 2-Kuru, 3-Nemli, 4-Islak, 5-Çok ıslak.

Bu sekiz farklı bitki türünün ıslaklık durumuna göre yayılışı araştırılmıştır. Bitkilerin yayılmalarının ortalama ıslaklıkları arasında anlamlı farklılık olup olmadığı test edilmiştir (Tablo 8).

Tablo 8 : Bitki Türleri ile Islaklık Durumuna Göre Tek Yönlü Varyans Analizi Bilgisayar Çıktısı

TÜR	Gözlem sayısı (n)	Ortalama (*)	Standart sapma	F-Değeri	Anlamlılık düzeyi (p)
Alnus glutinosa (1)	1362	3,7988	,8217	172,563	0,0001
Betula spp. (2)	822	3,0998	,7834		
Carpinus betulus (3)	857	3,0898	1,3413		
Fraxinus excelsior (4)	1058	3,6153	1,2366		
Ulmus spp. (5)	992	3,7359	1,4518		
Tilia europaea (6)	1411	3,0886	1,2337		
Pinus nigra (7)	1603	2,6419	1,0138		
Quercus petraea (8)	929	2,8052	1,3668		
Toplam	9034	3,2212	1,2372		

(*) Toprağın ıslaklık durumuna göre beşli ölçek kriterleri ortalama değeri

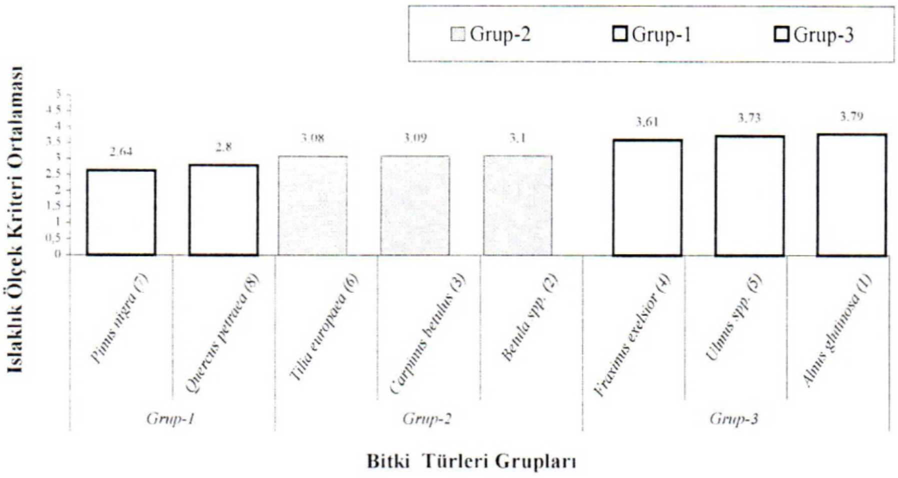
(1-Çok kuru, 2-Kuru, 3-Nemli, 4-Islak, 5-Çok ıslak)

P=0,0001*** < 0,001 olduğundan grup ortalamaları arasında ileri derecede anlamlı fark vardır. Buradaki anlamlılık oluşturan grupları tespit edebilmek için Scheffe alt-testi kullanılmıştır. Tablo-9'daki Scheffe alt testinde görüldüğü gibi farklı üç grup oluşmuştur.

Bu üç grubu ortalamaya göre küçükten büyüğe doğru sırasıyla yazarsak Grup-1: *Pinus nigra* (7), *Quercus petraea* (8); Grup -2: *Tilia europaea* (6), *Carpinus betulus* (3), *Betula* spp. (2); Grup -3: *Fraxinus excelsior* (4), *Ulmus* spp. (5), *Alnus glutinosa* (1).Böylece bu bitkilerden Grup-1 dekiler en az, Grup-2 dekiler orta, Grup-3 dekiler ise yüksek ıslaklık durumuna göre yayılma göstermektedir (POLUNIN/WALTERS 1985).

Tablo 9 : Scheffe Testine Göre Alt Gruplar

TÜR	Gözlem sayısı (n)	Anlamlılık düzeyi P = 0.05		
		Grup-1	Grup-2	Grup-3
Pinus nigra (7)	1603	2,6419	-	-
Quercus petraea (8)	929	2,8052	-	-
Tilia europaea (6)	1411	-	3,0886	-
Carpinus betulus (3)	857	-	3,0898	-
Betula spp. (2)	822	-	3,0998	-
Fraxinus excelsior (4)	1058	-	-	3,6153
Ulmus spp. (5)	992	-	-	3,7359
Alnus glutinosa (1)	1362	-	-	3,7988



Şekil 2 : Toprağın ıslaklık durumuna göre oluşan bitki türü grupları

KAYNAKLAR

AGNESTI, A., 1996 : Introduction to Categorical Data Analysis, John Wiley & Sons, New York, S.17.

BISHOP, Y.M.M., S.E.FİENBERG, P.W. HOLLAND, 1991: Discrete Multivariate Analysis: Theory and Practice, Eleventh Edition, Mit Press, Cambridge, S.374.

ERSOY, N., ORAL, S.E., 1992 : Olasılık ve İstatistiğe Giriş, Gazi Üniversitesi , Ankara, S.511.

GARSON, G.D., 1999 : Measures of Association,North Carolina State University Websites, http://www2.chass.ncsu.edu/garson/pa765/garson.htm, 1999.

KALIPSIZ, A., 1994: İstatistik Yöntemler, İ.Ü. Orman Fakültesi, İstanbul, S.558.

NIE, N., G. JENKINS, J., HULL, C. H., STEINBRENNER, K. DALE, B., 1975: SPSS (Statistical package for the social science). –MacGraw-Hill Book Company, New York.

POLUNIN, O. & WALTERS, M., 1985 : A Guide to the Vegetation of Britain and Europe, Oxford University Press, London.

POLUNİN, O. 1969 : Flowers of Europe, Oxford University Press, London.

SERPER, Ö. 1993 : Uygulamalı İstatistik 2, Filiz Kitabevi, İstanbul, S.356.

YOĞURTÇUGİL, K., 1978: Ki-Kare Üzerine Bir Deneme, İstanbul Üniversitesi İktisat Fakültesi, İstanbul, S.95.