

Fen Fakültesi İstatistik ve Bilgisayar Bilimleri Bölümü

# Çok Değişkenli İstatistiksel Yöntemler

Çok Değişkenli Hipotez Testleri

Dr. Öğr. Üyesi Uğur ŞEVİK



### Ders İçeriği

- BÖLÜM 4 ÇOK DEĞİŞKENLİ HİPOTEZ TESTLERİ
  - Çok Değişkenli Evren Ortalaması Önemlilik Testi.
  - Bağımsız İki Gruba İlişkin Ortalama Vektörlerinin Karşılaştırılması (Hotelling T²)



#### **Ders Hedefleri**

• Çok değişkenli analizdeki hipotez testlerinin nasıl yapıldığı öğrenilecektir.



- Bilindiği gibi parametrik hipotez testlerinin örneklemin çekildiği evren ile ilgili iki temel varsayımı vardır. Bunlar normallik ve varyansların homojenliği varsayımıdır.
- Bu varsayımların çok değişken için genelleştirilmiş durumu olan; çok değişkenli normallik ve varyans-kovaryans matrislerinin eşitliği varsayımları burada söz konusudur.
- Tek değişkenli testler, değişkenler arasındaki ilişkinin bir göstergesi olan kovaryansı (ya da korelasyonu) dikkate almadıklarından, değişkenler arasında var olan bilginin daha az bir bölümünü kullanırlar.
- Diğer taraftan, çok değişkenli hipotez testlerine ilişkin test istatistiklerinin hesaplanmasında, değişkenler arasındaki ilişkiler kovaryans matrisi aracılığı ile dikkate alınır.

- Birden çok değişken açısından gruplar arasında fark olup olmadığı tek değişkenli testlerle araştırıldığında, değişkenler açısından gruplar arasında anlamlı bir fark çıkmayabilir;
- Ancak çok değişkenli testlerle birleşik değişkenler kümesi arasında fark arandığından gruplar arasında fark çıkabilir.
- Diğer bir deyişle, birkaç değişkendeki küçük farklılıklar tümel farklılık yaratabilir.
- Çok değişkenli testler bu tür durumlarda daha güçlü olabilmektedir.
- Bu bölümde, çok kullanılan bazı tek değişkenli hipotez testlerinin çok değişkenli karşılıkları üzerinde durulacak ve bazı çok değişkenli testlerin tek değişkenli yaklaşımla elde edilen sonuçlarına da yer verilecektir.

- 1.Çok Değişkenli Evren Ortalaması Önemlilik Testi.
- 2.Birden Fazla Gruba İlişkin İki Eş Arasındaki Farkın Önemlilik Testinin Çok Değişkenli Evren Ortalaması Önemlilik Testi Yardımıyla Çözümü.
- 3.Varyans-Kovaryans Matrislerinin Eşitliğinin Test Edilmesi (Box-M testi).
- 4.Bağımsız İki Gruba İlişkin Ortalama Vektörlerinin Karşılaştırılması.
- 5.Çok Değişkenli Tek Yönlü Varyans Analizi (MANOVA).
- 6.Çok Değişkenli İki Yönlü Varyans Analizi.
- 7.Tekrarlı Ölçümlerde Tek Yönlü Varyans Analizi.
- 8.Tek Etken Üzerinde Tekrarlamaların Olduğu Tekrarlı Ölçümlerde İki Yönlü Varyans Analizi.
- 9.Korelasyon matrisinin birim matris olup olmadığının test edilmesi (Bartlett testi (Bartlett's test of sphericity).



#### 1. Çok Değişkenli Evren Ortalaması Önemlilik Testi.

- Bilindiği gibi tek değişkenli hipotez testlerinden "evren ortalaması önemlilik testi", örneklemden elde edilen bir ortalamanın, evren ortalaması olarak kabul edilen hipotetik bir değerden farklı olup olmadığını test etmekte kullanılır.
- Örneğin Doğu Anadolu Bölgesi'nde hemoglobin ortalamasının μ<sub>0</sub>=13,5 mg olup olmadığını öğrenilmek istendiğinde tek örneklem testlerinden evren ortalaması önemlilik testinden yararlanılır.
- Bu amaçla, belirli bir örneklem üzerinde çalışılarak elde edilecek örneklem ortalaması ve standart hatası sırasıyla  $(\bar{X})$  ve  $(S_{\bar{X}})$  ile gösterildiğinde, test istatistiği;

 $S_{\bar{x}} = \frac{1}{S/\sqrt{n}}$ 

• ile verilir.



#### 1. Çok Değişkenli Evren Ortalaması Önemlilik Testi.

- Yukarıda verilen test istatistiği n-1 serbestlik dereceli bir t dağılımı gösterir.
- Bu teste ilişkin H<sub>0</sub> ve H<sub>1</sub> hipotezlerinin aşağıdaki gibi iki yönlü kurulduğunu düşünelim;

 $H_0: \mu = \mu_0$   $H_1: \mu \neq \mu_0$ 

• Buna göre, hesapla bulunan t değerinin mutlak değeri, n-1 serbestlik dereceli ve  $\alpha/2$  yanılma düzeyindeki t tablo istatistiği  $(t_{(n-1;\alpha/2)})$ ile karşılaştırıldığında,

$$t_{Hesap} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} > t_{Tablo(n-1; \alpha/2)}$$

ise H<sub>0</sub> hipotezi reddedilir ve dolayısıyla H<sub>1</sub> hipotezi kabul edilir.

#### 1. Çok Değişkenli Evren Ortalaması Önemlilik Testi.

- Çok değişkenli evren ortalaması önemlilik testinde ise birden çok değişkene ilişkin örneklem ortalamasını içeren ortalama vektörü, evren ortalama vektörü ile karşılaştırılır.
- Diğer bir deyişle, çok değişkenli evren ortalaması önemlilik testi, tek değişkenli evren ortalaması önemlilik testinin birden çok değişken için genelleştirilmiş durumudur.
- H<sub>0</sub> ve H<sub>1</sub> hipotezleri aşağıdaki gibi kurulabilir;

$$H_{0}:\begin{bmatrix} \mu_{1} \\ \mu_{2} \\ \vdots \\ \mu_{p} \end{bmatrix} = \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0p} \end{bmatrix} \quad H_{1}: \begin{bmatrix} \mu_{1} \\ \mu_{2} \\ \vdots \\ \mu_{p} \end{bmatrix} \neq \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0p} \end{bmatrix}$$

- 1. Çok Değişkenli Evren Ortalaması Önemlilik Testi.
- Karara varabilmek için çok değişkenli T² istatistiği elde edildikten sonra F istatistiğine dönüşüm yapılır.
- T<sup>2</sup> istatistiğini bulabilmek için t istatistiği ifadesinin karesi alınırsa;

$$t^{2} = \frac{(\bar{x} - \mu_{0})^{2}}{s^{2}/n} = n(\bar{x} - \mu_{0})(s^{2})^{-1}(\bar{x} - \mu_{0}) > t_{(n-1;\alpha/2)}^{2}$$

yazılabilir. Bu ifade çok değişken için genellenirse;

 $\overline{x}$ : Örneklem ortalama vektörü

 $S^{-1}$ : Tekil olmayan örneklem varyans-kovaryans matrisi

μ<sub>0</sub>: Evren ortalama vektörü

olmak üzere T2;

$$T^2 = n(\bar{x} - \mu_0)' S^{-1}(\bar{x} - \mu_0)$$

olarak yazılır ve T2 istatistiği;

$$F = \frac{n-p}{(n-1)p}T^2$$

olmak üzere p ve n-p serbestlik dereceli bir F dağılımı gösterir.

- 1. Çok Değişkenli Evren Ortalaması Önemlilik Testi.
- Karar
- Aşağıda bulunan F hesap değeri p ve n-p serbestlik dereceli bir F tablo değerinden büyük ise H<sub>1</sub> hipotezi kabul edilir ve en az bir değişkenin ortalaması, ilgili evren ortalamasından farklıdır denir.

 $\bar{x}$ : Örneklem ortalama vektörü

 $S^{-1}$ : Tekil olmayan örneklem varyans-kovaryans matrisi

μ<sub>0</sub>: Evren ortalama vektörü

olmak üzere T2;

$$T^{2} = n(\bar{x} - \mu_{0})' S^{-1}(\bar{x} - \mu_{0})$$

olarak yazılır ve T2 istatistiği;

$$F = \frac{n-p}{(n-1)p}T^2$$

olmak üzere p ve n-p serbestlik dereceli bir F dağılımı gösterir.

#### 1. Çok Değişkenli Evren Ortalaması Önemlilik Testi.

#### Örnek:

 En az 20 kez milli olmuş, 18 yaşın üzerindeki erkek orta mesafe yüzücülerinin oksijen tüketim kapasitelerinin (MAXVO2) 68 ml/kg/dk, benç presteki maksimum kaldırma kuvvetlerinin (MKK) 40 kg ve genel spor bilgisi puanlarının (SBİL) 65 puan olup olmadığı öğrenilmek isteniyor. Bu amaçla belirlenen özelliklere uyanlar arasından 25 yüzücü rastgele seçilerek gerekli ölçümler yapılıyor. Araştırma sonuçları yandaki tabloda verilmiştir.

Kişi No	MAXVO2	MKK	SBİL	
1	62	48	68	
2	68	42	75	
3	57	38	79	
4	56	40	78	
5	74	44	83	
6	62	48	66	
7	62	41	80	
8	52	33	75	
9	63	36	75	
10 73		43	73	
11	65	43	74	
12	55	37	65	
13	57	47	62	
14	60	32	64	
15	63	45	70	
16	61	48	74	
17	67	53	76	
18	67	46		
19	62	42	76	
20	63	43 84		
21	55	37	74	
22	58	42	73	
23	66	43	74	
24	58	36	67	
25	64	42	75	
Ortalama	62,00	41,96	73,68	
Varyans	29,333	25,457	34,310	



- 1. Çok Değişkenli Evren Ortalaması Önemlilik Testi.
- Çözüm:
- Tabloya ilişkin ortalama vektörü ( $\bar{X}$ ), varyans-kovaryans matrisi (S) ve varyans kovaryans matrisinin inversi (S-1) ile evren ortalama vektörü ( $\mu_0$ ), aşağıda verilmiştir.

$$\bar{x}' = \begin{bmatrix} 62,00 & 41,96 & 73,68 \end{bmatrix} \qquad \mu'_0 = \begin{bmatrix} 68,00 & 40,00 & 65,00 \end{bmatrix} \\
S = \begin{bmatrix} 29,333 & 13,333 & 11,583 \\ 13,333 & 25,457 & 2,612 \\ 11,583 & 2,612 & 34,310 \end{bmatrix} \\
S^{-1} = \begin{bmatrix} 0,05186 & -0,02556 & -0,01556 \\ -0,02556 & 0,05219 & 0,00466 \\ -0,01556 & 0,00466 & 0,03404 \end{bmatrix}$$

- 1. Çok Değişkenli Evren Ortalaması Önemlilik Testi.
- · Çözüm:
- Çözümlemeye geçmeden önce, verilerin çok değişkenli normal dağılıp dağılmadığının incelenmesi gerekir.

Normallik testleri bu veriye uygulandığında, normallikten aşırı bir ayrılış olmadığı görülecektir.

olmadığı görülecektir. 
$$T^2 = 25 \begin{bmatrix} 62,00 - 68,00 & 41,96 - 40,00 & 73,68 - 65,00 \end{bmatrix} \begin{bmatrix} S^{-1} \end{bmatrix} \begin{bmatrix} 62,00 - 68,00 \\ 41,96 - 40,00 \\ 73,68 - 65,00 \end{bmatrix}$$

$$T^2=175,3$$
 olarak bulunur. F hesap istatistiği ise;  $F = \frac{25-3}{(25-1)3}175,3 = 53,56$ 

- olarak bulunur ve  $F_{Hesap} = 53,56 > F_{Tablo\,(3,\,22;\,\alpha=0,05)} = 3,05$  olduğu için  $H_0$  hipotezi reddedilir.
- Sonuçta; örneklem ortalama vektöründeki en az bir değişkenin ortalaması ilgili evren ortalamasından farklıdır denir.
- Diğer bir deyişle  $H_0$ :  $\mu = \mu_0$  hipotezi reddedilerek, örneklemin varsayılan evrenden çekilmediği sonucuna varılır.

## 2. Bağımsız İki Gruba İlişkin Ortalama Vektörlerinin Karşılaştırılması (Hotelling T2)

- Tek değişkenli çözümlemede, bir değişken yönünden bağımsız iki gruba ilişkin ortalamaların karşılaştırılması, parametrik test varsayımlarının (normallik ve varyansların homojenliği) sağlanması durumunda "iki ortalama arasındaki farkın önemlilik testi" ile yapılıyordu.
- Örneğin;
- 1. Hasta ve sağlam gruplarında hemoglobin yönünden fark olup olmadığı
- 2.1-9 ve 10-15 yıldır spor yapan uzun mesafe koşucularının submaksimal yükleme sonrasındaki geri toparlanma süreleri arasında fark olup olmadığı
- 3. 0,5 doz ile 1,5 doz ilaç verilen haşerelerin ölüm süreleri arasında fark olup olmadığı şeklindeki çalışmalar iki ortalama arasındaki farkın önemlilik testi ile karşılaştırılıyordu.

# 2. Bağımsız İki Gruba İlişkin Ortalama Vektörlerinin Karşılaştırılması (Hotelling T2)

- Bu örneklerden ilk ikisi, aşağıdaki şekilde çok değişkenli konuma getirilebilir.
- Birinci örnekte; hasta ve sağlam gruplarındaki hemoglobin ile birlikte örneğin potasyum, kalsiyum gibi değişkenler de ele alınır ise, sorun; "hasta ve sağlam grupları arasında hemoglobin, potasyum, kalsiyum,.... değişkenleri açısından bir fark var mıdır?" şeklinde çok değişkenli bir soruna dönüşecektir (Tablo 8.3).

	HASTA SAĞLAM				
Hemoglobin	Potasyum	Kalsiyum	Hemoglobin	Potasyum	Kalsiyum
	g . William tris	v uno realis	BAGIRAN	A ab Jour I	PROJECT IN

## 2. Bağımsız İki Gruba İlişkin Ortalama Vektörlerinin Karşılaştırılması (Hotelling T²)

• Yine, ikinci örnekte, spor yaşı 1 -9 ve 10- 15 yıl olan gruplarda geri toparlanma süresinin (sn) yanında bu sporcuların MAXVO2'lerinin de farklı olup olmadığı öğrenilmek istendiğinde çok değişkenli bir sorun ile karşılaşılacaktır (Tablo 8.4).

melecularismulation and	SPOR	Y A Ş I (yıl)	to nin show	
1–9		10–15		
Geri Toparlanma Süresi (sn)	MAXVO2 (ml/kg/dk)	Geri Toparlanma Süresi (sn)	MAXVO2 (ml/kg/dk)	
41	52	32	67	
ne de mannets du pris	antidated attention	The state of the s	satisfa de o al	
35	56	28	65	

 Bu örneklerdeki sorun; bağımsız iki gruba ilişkin ortalama vektörlerinin karşılaştırılması sorunudur. Bağımsız iki gruba ilişkin ortalama vektörlerinin karşılaştırılmasında da T² (Hotelling's T²) istatistiğinden yararlanılır.



# 2. Bağımsız İki Gruba İlişkin Ortalama Vektörlerinin Karşılaştırılması (Hotelling T²)

- Varsayımlar:
- Çok değişkenli iki ortalama vektörünün karşılaştırılmasında, her iki gruptaki verilerin ayrı ayrı normal dağılım göstermesi ve bu iki gruba ilişkin kovaryans matrislerinin homojen olması varsayımları vardır.
- Genel olarak, çok değişkenli 2 grubun karşılaştırılmasında kullanılan T² ile
   2'den fazla çok değişkenli grubun karşılaştırılmasını sağlayan MANOVA, tek
   yönlü varyans analizinden (ANOVA) daha fazla gözleme gereksinim duyar.
- Bu nedenle bazı kaynaklarda, gruplardaki gözlem sayısının 20'den fazla olması özellikle belirtilmektedir.

- 2. Bağımsız İki Gruba İlişkin Ortalama Vektörlerinin Karşılaştırılması (Hotelling T²)
- Hipotezler ve Test İstatistiği
- H<sub>0</sub> hipotezi; μ<sub>1</sub> ve μ<sub>2</sub> ortalama vektörleri olmak üzere, bu iki ortalama vektörü arasında fark olmadığı şeklinde kurulur (H<sub>0</sub>: μ<sub>1</sub>= μ<sub>2</sub>). H<sub>1</sub> hipotezi ise H<sub>1</sub>: μ<sub>1</sub>≠ μ<sub>2</sub> şeklindedir. Yukarıdaki varsayımlar altında;

 $n_1$ : 1. grubun gözlem sayısı

 $n_2$ : 2. grubun gözlem sayısı

 $\bar{x}_1$ : 1. grubun ortalama vektörü

 $\overline{x}_2$ : 2. grubun ortalama vektörü

S: Ortak varyans-kovaryans matrisi

olmak üzere T² istatistiği;



# 2. Bağımsız İki Gruba İlişkin Ortalama Vektörlerinin Karşılaştırılması (Hotelling T²)

Hipotezler ve Test İstatistiği

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\overline{x}_1 - \overline{x}_2)' S^{-1} (\overline{x}_1 - \overline{x}_2)$$
 ile verilir. Eşitlik (8.16) ile verilen  $T^2$  istatistiği; 
$$F = \frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} T^2$$
 ile  $p$  ve  $n_1 + n_2 - p - 1$  serbestlik dereceli bir  $F$  dağılımı gösterir.

- Eğer yukarıda verilen F hesap istatistiği F tablo istatistiğinden büyükse bağımsız iki ortalama vektörü arasında fark olduğu söylenir.
- İki ortalama vektörü arasında fark bulunursa, farklılığın hangi değişken(ler)den kaynaklandığı araştırılır.
- Çok değişkenli çözümlemede, ortalama vektörleri arasında fark olup olmadığının araştırılmasında sıklıkla kullanılan dört test istatistiği; Wilks lamda istatistiği, Pillai iz istatistiği, Hotelling iz istatistiği ve Roy'un en büyük özdeğere bağlı test istatistiğidir.

## 2. Bağımsız İki Gruba İlişkin Ortalama Vektörlerinin Karşılaştırılması (Hotelling T2)

- Örnek:
- Hem koroner arter hastalığı hem de diabet (şeker hastalığı) tanısı konan 26 hasta ile sadece diabet tanısı konan 24 hastaya ilişkin kolesterol, şeker ve HDL (High Density Lipoprotein) değerleri aşağıda verilmiştir. Ölçümler mg/dl cinsindendir. Acaba bu üç değişken açısından hasta gruplan arasında (bu iki gruba ilişkin ortalama vektörleri arasında) fark var mıdır?
- Bu örnekteki hasta gruplarına ilişkin varyans-kovaryans matrisleri ile ortak varyans-kovaryans matrisi, "varyans-kovaryans matrislerinin homojenliği" konusu incelendiğinde varyans-kovaryans matrisleri eşittir.
- Ayrıca, bu grupların çok değişkenli normal dağılıp dağılmadıklarına bakıldığında normallikten aşırı bir sapma olmadığı görülecektir.

2. Bağımsız İki Gruba İlişkin Ortalama Vektörlerinin Karşılaştırılması

(Hotelling T2)

• Örnek:

Diabet + Koroner Arter Hastaları		Diabet Hastaları			
Kolesterol	Şeker	HDL	Kolesterol	Şeker	HDL
210	130	44	215	160	63
249	185	25	234	150	58
265	183	46	238	190	72
280	200	23	255	169	74
230	175	35	212	145	65
290	170	25	264	185	60
270	150	28	245	176	63
260	188	38	220	168	57
295	172	39	268	185	62
218	165	32	220	175	67
222	165	30	224	180	50
235	167	43	207	176	62
240	178	36	230	170	54
242	180	37	217	151	67
270	169	47	245	166	62
285	180	39	257	175	66
275	160	48	249	173	60
269	145	40	244	190	56
255	140	36	235	139	63
259	190	35	236	118	66
265	125	50	237	163	64
268	130	31	250	152	68
270	120	47	244	196	58
274	190	37	275	205	67
310	215	38			
295	178	27			
rt: 261,577	167,308	36,769	238,375	169,042	62,667
/ar: 633,214	576,862	58,345	329,201	387,172	30,493

# 2. Bağımsız İki Gruba İlişkin Ortalama Vektörlerinin Karşılaştırılması (Hotelling T2)

· Örnek:

$$\bar{x}_1 = \begin{bmatrix} 261,577 \\ 167,308 \\ 36,769 \end{bmatrix} \qquad \bar{x}_2 = \begin{bmatrix} 238,375 \\ 169,042 \\ 62,667 \end{bmatrix} \qquad S^{-1} = \begin{bmatrix} 0,002345 & -0,00086 & -0,00101 \\ -0,00086 & 0,002558 & 0,002465 \\ -0,00101 & 0,002465 & 0,024611 \end{bmatrix}$$

Eşitlik (8.16)'dan,

$$T^{2} = \frac{(26)(24)}{26+24} \left[ 261577 - 238375 \quad 167,308 - 169,042 \quad 36,769 - 62,667 \right] S^{-1} \begin{bmatrix} 261,577 - 238,375 \\ 167,308 - 169,042 \\ 36,769 - 62,667 \end{bmatrix}$$

$$T^2 = 12,48 \begin{bmatrix} 23,202 & -1,734 & -25,898 \end{bmatrix} S^{-1} \begin{bmatrix} 23,202 \\ -1,734 \\ -25,898 \end{bmatrix}$$

$$T^2 = 12,48(19,28122) = 240,63$$



- 2. Bağımsız İki Gruba İlişkin Ortalama Vektörlerinin Karşılaştırılması (Hotelling T2)
- · Örnek:

$$F = \frac{26 + 24 - 3 - 1}{(26 + 24 - 2)(3)} 240,63 = 76,868$$

olarak bulunur.  $F_{Hesap} = 76,868 > F_{Tablo\,(3,\,46;\,\alpha=0,05)} \cong 2,84$  olduğu için iki ortalama vektörü arasında fark olduğu sonucuna ulaşılır. Ortalama vektörleri arasında fark olduğu için farklılığın hangi değişken(ler) arasında olduğunun araştırılması gerekir. (not: İstatistiksel yazılımlarda bu sonuç 76,848 olarak bulunmaktadır. Bu durum, kitaptaki çözümlemelerin elementer matris işlemleri ile yapılmasından kaynaklanmaktadır.)

### ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ

Sunum hazırlanırken aşağıdaki kaynaktan yararlanılmıştır.





