



T.C.
ONDOKUZ MAYIS ÜNİVERSİTESİ

FEN EDEBİYAT FAKÜLTESİ
İSTATİSTİK BÖLÜMÜ

İST.482 PARAMETRİK OLMAYAN
İSTATİSTİKSEL YÖNTEMLER

PROF. DR. YÜKSEL ÖNER

6. Hafta

4.4 RASTGELELİK İÇİN DİZİ PARÇALARI (RUN) TESTİ

İstatistiksel verilerin analizinde sıklıkla karşılaşılan kavramlardan birisi rastgeleliktir. İstatistiksel testlerin hemen hemen hepsinde ilgili kitleden örnek çekimi rastgelelik esasına dayandırılır. Gözlem değerlerine ait hata terimlerinin bir rastgele değişken olduğu kabul edilir.

Tek kitle veya grupta rastgelelik için dizi parçaları testi, bir örnekten elde edilen gözlem değerlerinin belirlenen bir kritere göre dağılımlarının rastgele olup olmadığını test etmek için kullanılmaktadır. Burada kriter gözlem birimlerinin iki tür simge ile gösterilmesi olabileceği gibi, belli bir değerden (ortalama veya medyan gibi) sapmalar da olabilir. Test işleminin algoritması şu şekilde verilebilir.

1. Hipotezler kurulur

H_0 : Simgelerin dizilişi rastgeledir (veya belirlenen bir değerden sapmaların dizilişi rastgeledir)

H_1 : Simgelerin dizilişi rastgele değildir (veya belirlenen bir değerden sapmaların dizilişi rastgele değildir)

2. İlgili kitleden n birimlik örnek çekilerek her bir örnek birimi sırasıyla iki tür simgeden (A veya B; + veya -; Başarılı veya başarısız; ortalamanın altında veya ortalamanın üstünde v.s.) birisi ile simgelenir. Böyle iki türdeki simgelerin bir sıralı dizisi elde edilir. Bu simgeler dizisinde aynı türden simgelerin oluşturdukları gruplara “dizi parçası” adı verilir. Bir dizi parçasında bir tek simge olabileceği gibi, birden fazla da simge bulunabilir.

3. Test istatistiği belirlenir. Run testinde test istatistiği;

r : Dizi parçalarının sayısı

şeklinde tanımlı olup alabileceği değerler $r = 2, 3, 4, 5, \dots, n_1 + n_2$ dir. Burada n_1 birinci tür simgelerin sayısı ve n_2 ikinci tür simgelerin sayısı olmak üzere $n = n_1 + n_2$ dir. Test istatistiğinin örnekten hesaplanan değeri r_h ile gösterilsin.

4. Karar kuralı belirlenir ve karar verilerek sonuç değerlendirilir.

H_0 hipotezi doğru iken, yani simgelerin dizilişinde bir rastgelelik varken test istatistiğinin yeteri kadar küçük veya yeteri kadar büyük değerler almaması beklenir. Diğer bir ifadeyle simgelerin dizilişinde kümelenme (aynı türden simgelerin bir arada bulunması) durumunun veya sistematik (bir birinci türden bir ikinci türden simgenin ardışık olarak) diziliş göstermesi durumunun ortaya çıkması beklenmez. Çünkü bu diziliş durumlarında rastgelelik olmayıp, birinci durumda test istatistiği en küçük değerini alacak ve $Enk(r) = 2$ olacaktır. İkinci durumda ise test istatistiği en büyük değerini alacak ve $Enb(r) = n_1 + n_2 = n$ olacaktır. Bu durumlar H_1 hipotezini destekleyecektir. Buna göre H_1 hipotezi doğru iken test istatistiğinin ya çok küçük bir değer ya da çok büyük bir değer alması beklenir. Ne kadar küçük veya ne kadar büyük olması gerektiğini belirleyecek olan kritik değerler sırası ile α önem seviyesi için $r_{\frac{\alpha}{2}}$ ve

$r'_{\frac{\alpha}{2}}$ olmak üzere, eğer $r_h \leq r_{\frac{\alpha}{2}}$ veya $r_h \geq r'_{\frac{\alpha}{2}}$ ise H_0 hipotezi ret edilir. $r_{\frac{\alpha}{2}} \leq r_h \leq r'_{\frac{\alpha}{2}}$ ise H_0

hipotezi ret edilemez. $r_{\frac{\alpha}{2}}$ kritik değeri $\alpha = 0,05$ iken n_1 ile n_2 dikkate alınarak T11(a) tablosundan ve $r'_{\frac{\alpha}{2}}$ kritik değeri de yine $\alpha = 0,05$ iken n_1 ile n_2 dikkate alınarak T11(b) tablosundan belirlenir.

Sonuç olarak H_0 hipotezi ret edilirse simgelerin dizilişi rastgele değildir, H_0 hipotezi ret edilemezse simgelerin dizilişi rastgeledir.

Örnek 4.13 Bir kuyrukta yer alan 8 bireyin cinsiyete göre farklı diziliş biçimleri aşağıda verilmiştir. Bu dizilişlerin her birinde dizi parçalarını belirleyerek test istatistiğinin alabileceği değerleri bulunuz?

a)

Cinsiyet: K K E E K E E K	$r_h = 5$
-----------------------------------------------------------------------------------	-----------

b)

Cinsiyet: E E E E K K K K (Kümelene Durumu)	$r_h = 2$
--------------------------------------------------------------------------------------------------------	-----------

c)

Cinsiyet: E K E K E K E K (Sistemik diziliş durumu)	$r_h = 8$
----------------------------------------------------------------------------------------------------------------	-----------

d)

Cinsiyet: E K K E K E E K	$r_h = 6$
-----------------------------------------------------------------------------------	-----------

4.4.1 SPSS’de Rastgelelik için Dizi Parçaları Testi

Spss’de rastgelelik için dizi parçaları testini uygulanmasında takip edilecek algoritmanın adımları şu şekilde verilebilir.

1.Adım Variable View sayfasında değişkenler ve özelliklerinin tanımlamaları yapılır. **Data View** sayfasında değişkenlere ait veriler girilir.

2.Adım Analyze > Nonparametric Tests > Legacy Dialogs > Runs... yolu izlenerek açılacak olan ekranda değişkenler listesinden ilgilenilen değişken seçilerek **Test Variable List** işlem kutusuna aktarılır. Bu ekranda yer alan **Cut Point** seçenekler kutusundan hangi kritere göre birimlerin simgelenmesi yapılacaksa o kriter seçilir. Burada **Median** “Medyan”, **Mean** “Ortalama”, **Mod** “Mod (en çok tekrarlanan gözlem) değeri”, **Custom** “belirlenen her hangi bir değer” anlamındadır.

3.Adım Exact... penceresi açılarak **Exact** seçeneği işaretlenir. **Continue** ve **Ok** tıklanarak işlem bitirilir. Sonuç tablosu çıktı sayfasında sunulur.

4.Adım α önem seviyesinde H_1 hipotezine göre karar kuralı belirlenir ve H_0 hipotezi hakkında karar verilir. H_1 hipotezi çift yönlü olduğundan sonuç özet tablosunda bulunan **Exact Sig. (2-sided test)** “**Kesin önemlilik olasılığı (iki yönlü test için)**” ifadesinin karşısındaki sayısal değer, önemlilik olasılığı olarak bilinen p olasılığının iki katıdır. Bu durumda $p = (\text{sayısal değer}/2)$ olarak alınır ve $\frac{\alpha}{2}$ önem seviyesi ile karşılaştırılır. $p = (\text{sayısal değer}/2)$ olmak üzere $p < \frac{\alpha}{2}$ ise H_0 hipotezi ret edilir, $p \geq \frac{\alpha}{2}$ ise H_0 hipotezi ret edilemez.

Örnek 4.14 Bir kitleden çekilen 30 birimlik bir örnekte örnek birimlerinin bir X değişkeni üzerinde ölçüm değerleri aşağıda verilmiştir. Önem seviyesi %5 olmak üzere

- a) Ortalamadan sapmaların rastgelelik gösterip göstermediğini inceleyiniz?
b) Medyan değerinden sapmaların rastgelelik gösterip göstermediğini inceleyiniz?

X_i					
105	92	98	65	133	160
108	96	115	69	98	162
116	142	122	72	96	154
119	145	136	141	122	91
80	96	127	146	129	88

Cözüm: a) Test edilecek olan hipotezler:

H_0 : Ortalamadan sapmaların dizilişi rastgeledir

H_1 : Ortalamadan sapmaların dizilişi rastgele değildir

Test istatistiği ve test istatistiğinin örnekte alabileceği değer; test istatistiği

r : Dizi parçalarının sayısı

olup, örnekte alabileceği değeri bulabilmek için verilen veriyi önce ortalamadan sapmalar dizisine dönüştürelim. Bunun için ortalamanın altında olan gözlem değerlerini “A” ve ortalamanın üstünde olan gözlem değerlerini ise “B” ile simgeleyelim. Verilen örneğin örnek ortalaması $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{3423}{30} = 114,1$ bulunur.

Sıra No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X_i	105	108	116	119	80	92	96	142	145	96	98	115	122	136	127
Simge	A	A	B	B	A	A	A	B	B	A	A	B	B	B	B

Sıra No	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
X_i	65	69	72	141	146	133	98	96	122	129	160	162	154	91	88
Simge	A	A	A	B	B	B	A	A	B	B	B	B	B	A	A

$n_1 = 14$ (A'ların (birinci tür simgelerin) sayısı), $n_2 = 16$ (B'lerin (ikinci tür simgelerin) sayısı), $n = 30$ ve dizi parçalarının sayısı $r_h = 11$ dir.

Karar: $\alpha = 0,05$ için kritik değerler $n_1 = 14$ ve $n_2 = 16$ iken T11(a) tablosundan $r_{\frac{\alpha}{2}} = 10$ ve T11(b) tablosundan $r'_{\frac{\alpha}{2}} = 22$ elde edilir. Bu sonuçlara göre $10 < 11 < 22$ yani $r_{\frac{\alpha}{2}} < r_h < r'_{\frac{\alpha}{2}}$ olduğundan H_0 hipotezi ret edilemez. Yani ortalamadan sapmaların dizilişi rastgeledir.

b) Test edilecek olan hipotezler:

H_0 : Medyandan sapmaların dizilişi rastgeledir

H_1 : Medyandan sapmaların dizilişi rastgele değildir

Test istatistiği ve test istatistiğinin örnekte alabileceği değer; test istatistiği

r : Dizi parçalarının sayısı

olup, örnekte alabileceği değeri bulabilmek için verilen veriyi önce medyandan sapmalar dizisine dönüştürelim. Verilen örneğin örnek medyanı $n = 30$ çift sayı olduğundan, $Md = \frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2} = \frac{X_{(15)} + X_{(16)}}{2} = \frac{115 + 116}{2} = 115,5$ bulunur. Buna göre 115,5'dan küçük olan gözlemleri "A" ve 115,5'dan büyük olan gözlemleri de "B" ile simgeleyelim.

Sıra No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X_i	105	108	116	119	80	92	96	142	145	96	98	115	122	136	127
Simge	A	A	B	B	A	A	A	B	B	A	A	A	B	B	B

Sıra No	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
X_i	65	69	72	141	146	133	98	96	122	129	160	162	154	91	88
Simge	A	A	A	B	B	B	A	A	B	B	B	B	B	A	A

$n_1 = 15$ (A'ların (birinci tür simgelerin) sayısı), $n_2 = 15$ (B'lerin (ikinci tür simgelerin) sayısı), $n = 30$ ve dizi parçalarının sayısı $r_h = 11$ dir.

Karar: $\alpha = 0,05$ için kritik değerler $n_1 = 15$ ve $n_2 = 15$ iken T11(a) tablosundan $r_{\frac{\alpha}{2}} = 10$ ve T11(b) tablosundan $r'_{\frac{\alpha}{2}} = 22$ elde edilir. Bu sonuçlara göre $10 < 11 < 22$ yani $r_{\frac{\alpha}{2}} < r_h < r'_{\frac{\alpha}{2}}$ olduğundan H_0 hipotezi ret edilemez. Yani medyandan sapmaların dizilişi rastgeledir.

SPSS çözümü

<p>a) H_0: Ortalamadan sapmaların dizilişi rastgeledir H_1: Ortalamadan sapmaların dizilişi rastgele değildir</p>	<p>b) H_0: Medyandan sapmaların dizilişi rastgeledir H_1: Medyandan sapmaların dizilişi rastgele değildir</p>
---------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------

a)		b)	
Runs Test 1		Runs Test 2	
	x		X
Test Value ^a	114,1000	Test Value ^a	115,50
Cases < Test Value	14	Cases < Test Value	15
Cases >= Test Value	16	Cases >= Test Value	15
Total Cases	30	Total Cases	30
Number of Runs	11	Number of Runs	11
Z	-1,655	Z	-1,672
Asymp. Sig. (2-tailed)	,098	Asymp. Sig. (2-tailed)	,094
Exact Sig. (2-tailed)	,090	Exact Sig. (2-tailed)	,091
Point Probability	,027	Point Probability	,026

a. Mean

$p = \frac{0,090}{2} = 0,045$ ve $\alpha = 0,05$ iken $p > \frac{\alpha}{2} = 0,025$ olduğundan H_0 hipotezi ret edilemez. Ortalamadan sapmaların dizilişinin rastgele olduğu %95 güvenle söylenebilir.

b. Median

$p = \frac{0,091}{2} = 0,045$ ve $\alpha = 0,05$ iken $p > \frac{\alpha}{2} = 0,025$ olduğundan H_0 hipotezi ret edilemez. Medyandan sapmaların dizilişinin rastgele olduğu %95 güvenle söylenebilir.

SORULAR

SORU:1 Bir tarımcı tohumları aldığı firmanın görüşünün tersine olarak tohumlarda çimlenenlerin oranının yaşadığı deneyimlere göre %80 den küçük olduğunu düşünmektedir. İlgili firmadan alınan tohumlardan 9 tanesi rastgele seçilmiş ve yapılan ekim sonucunda Ç: çimlendi, N: çimlenmedi olmak üzere aşağıdaki veri düzeni elde edilmiştir. Buna göre tarımcının düşüncesinde haklı olup olmadığına %10 önem seviyesinde karar veriniz?

Tohum No	1	2	3	4	5	6	7	8	9
Çimlenme durumu	Ç	N	Ç	N	Ç	Ç	Ç	N	Ç

SORU:2 Yabancı dil kursu veren bir dersane yabancı dil sınavında 70 barajını geçen öğrenci oranının 0.80'den fazla olduğunu ileri sürmektedir. Yabancı dil kursuna devam eden öğrencilerden 7 tanesi rastgele seçilmiş ve sınavda aldıkları puanlar aşağıdaki gibi gözlenmiştir. Bu sonuçlara göre dershanenin iddiasında haklı olup olmadığına %10 önem seviyesinde karar veriniz?

Öğrenci no	1	2	3	4	5	6	7
Not	62	80	75	78	84	85	50

SORU:3 Belli bir hasta grubunda kandaki tahlil sonucu sınır değerin üstünde olanların oranının %70 den fazla olduğu iddia edilmektedir. Bu tahlil için sınır değer 110 birimdir. Bu hastaların içinden rastgele seçilen 8 hastanın tahlil sonuçları aşağıda verilmiştir. Buna göre iddianın doğruluğunu %5 önem seviyesinde kontrol ediniz?

Hasta no	1	2	3	4	5	6	7	8
Tahlil sonucu	105	112	120	100	125	128	130	195

SORU:4 $H_0: \pi = 0,60$ hipotezini $H_1: \pi \neq 0,60$ hipotezine karşı Binom testi ile test etmek için ilgili kitleden 8 birimlik bir örneklem çekilmiştir. Buna göre:

- Test istatistiğinin örnekleme dağılımını oluşturunuz?
- Test istatistiğinin beklenen değer ve varyansının hesaplayınız?
- %10 önem seviyesinde test istatistiğinin hangi değerleri için H_0 hipotezi ret edilebilir belirleyiniz?

SORU:5 $H_0: M = 30$ hipotezini, $H_1: M > 30$ hipotezine karşı işaret test ile test etmek için medyanı M olan kitleden rastgele olarak 42 birimlik bir örneklem çekilmiş ve örnek birimleri üzerinde $D_i = X_i - 30$, $i = 1, 2, \dots, 42$ dönüşümü uygulanmıştır. Bu dönüşüm sonucunda 6 tanesi için $D_i < 0$; 2 tanesi için $D_i = 0$ ve geriye kalanlar için $D_i > 0$ olduğu gözlenmiştir. Bu sonuçlara göre %5 önem seviyesinde H_0 hipotezi hakkındaki kararınızı belirtiniz?

SORU:6 Bir sektörde aylık gelire ait medyan parametresinin 1200 TL den fazla olduğu iddia edilmektedir. Bu sektörde çalışanlar arasından rastgele seçilen 12 çalışana ait aylık gelirler (TL) aşağıdaki gibi gözlenmiştir. Bu örneklemin söz konusu iddiayı destekleyip desteklemediğine %5 önem seviyesinde Wilcoxon işaretli sıra sayıları testine göre karar veriniz?

$X_i(TL): 1210 \ 1196 \ 1195 \ 1220 \ 1260 \ 1245 \ 1211 \ 1244 \ 1201 \ 1279 \ 1295 \ 1310$

SORU:7 A ilacının etki süresine ait medyanın 10 dakikadan farklı olup olmadığı test edilmek isteniyor. Rastgele seçilen 10 hastaya A ilacı uygulanmış ve her bir hasta için etki süresi aşağıdaki gibi ölçülmüştür. Hipotezleri oluşturarak, H_0 hipotezi hakkındaki kararınızı %10 önem seviyesinde

- İşaret testini kullanarak
- Wilcoxon işaretli sıra sayıları testini kullanarak, belirtiniz?

$X_i(Dakika): 13 \ 10 \ 7 \ 7 \ 12 \ 16 \ 17 \ 17 \ 19 \ 20$

SORU:8 $H_0: M = M_0$ hipotezini $H_1: M \neq M_0$ hipotezine karşı İşaret testi ile test etmek için medyanı M olan kitleden 10 birimlik bir örneklem çekildiğini ve bütün örnek birimlerine ait $D_i = X_i - M_0$ farkları için $D_i \neq 0$ olduğunu kabul edelim. Buna göre:

- Test istatistiğinin örnekleme dağılımını oluşturunuz?
- Test istatistiğinin beklenen değer ve varyansının hesaplayınız?
- %5 önem seviyesinde test istatistiğinin hangi değerleri için H_0 hipotezi kabul edilebilir belirleyiniz?

SORU:9 $H_0: M = M_0$ hipotezini $H_1: M > M_0$ hipotezine karşı Wilcoxon İşaretli Sıra Sayıları testi ile test etmek için medyanı M olan kitleden 3 birimlik bir örneklem çekildiğini, bütün örnek birimlerine ait $D_i = X_i - M_0$ farkları için $D_i \neq 0$ ve $|D_i|$ lerin birbirinden farklı olduğunu kabul edelim. Buna göre:

a) Test istatistiğinin örnekleme dağılımını oluşturunuz?

b) Test istatistiğinin beklenen değer ve varyansının hesaplayınız?

SORU:10 Bir firma ürettiği ürünlerin sadece %10'unun hatalı olduğunu iddia etmektedir. Firmanın ürettiği ürünler arasından her biri 5'er birimlik 100 numune seçilmiş ve hatalı ürün sayısına göre bu numunelerin dağılımı aşağıdaki gibi elde edilmiştir. Firmanın üretimi ile ilgili olarak, hatalı ürün sayısının Binom dağılımı ile uyumlu olup olmadığına %1 önem seviyesinde Ki-Kare uyum iyiliği testi ile karar veriniz?

Hatalı ürün sayısı	Numune sayısı
0	30
1	18
2	15
3	10
4	15
5	12

SORU:11 İkinci soru için yabancı dil sınav notlarına ait medyanın 70'den büyük olup olmadığına %5 önem seviyesinde a) işaret testi ile b) Wilcoxon işaretli sıra sayıları testi ile karar veriniz?

SORU:12 Bir işletme ürettiği cep telefonlarının ağırlıklarının 100 gram olmasını hedeflemektedir. Bu üretim sürecinden rastgele seçilen 20 cep telefonuna ait ağırlıklar aşağıdaki gibi ölçülmüştür. Bu örnekleme göre;

a) Ortalama değerden sapmaların b) Medyan değerinden sapmaların c) Hedeflenen değerden sapmaların

rastgelelik gösterip göstermediğine %5 önem seviyesinde karar veriniz?

Ağırlıklar				
99.09	98.14	99.81	101.43	98.48
99.96	98.23	99.68	99.13	101.01
99.72	99.68	99.66	99.55	100.74
99.64	101.63	98.84	98.21	100.58

SORU:13 Biyoloji, Kimya ve Fizik bölümlerinde okuyan öğrencilerden 7'şer tanesi rastgele seçilmiş ve Matematik dersinden aldıkları notlar aşağıdaki gibi elde edilmiştir. %5 önem seviyesinde;

a) Bu örneklerin her birinin normal dağılım ile uyumlu olup olmadığını Kolmogorov-Smirnov testi ile araştırınız?

b) Brown-Forsythe testini kullanarak varyansların homojenliğini test ediniz?

c) Levene testini kullanarak varyansların homojenliğini test ediniz?

BÖLÜM		
Biyoloji	Kimya	Fizik
40	52	70
50	55	72
42	63	80
54	60	80
60	82	90
57	61	92
37	65	80

SORU:14 Bir üniversitede okuyan öğrencilerin 400 tanesi rastgele seçilmiş ve günlük harcamalarına göre dağılımı aşağıdaki gibi elde edilmiştir. Günlük harcamalarına göre öğrenciler kitlesinin dağılımının $N(\mu = 16, \sigma^2 = 44)$ dağılımı ile uyumlu olup olmadığına %5 önem seviyesinde Ki-Kare uyum iyiliği testine göre karar veriniz?

Harcama(TL)	Öğrenci sayısı
$0 \leq X < 5$	30
$5 \leq X < 10$	50
$10 \leq X < 15$	100
$15 \leq X < 20$	120
$20 \leq X < 25$	80
$25 \leq X < 30$	12
$30 \leq X < 35$	8

SORU:15 Bir sınıftan rastgele seçilen 25 öğrencinin İstatistik dersinden almış oldukları notlar aşağıdadır. Bu örneklemin dağılımının

a) $N(\mu = 50, \sigma^2 = 144)$ dağılımı ile uyumlu olup olmadığına Kolmogorov-Smirnov testi ile

b) $N(\mu, \sigma^2)$ dağılımı ile uyumlu olup olmadığına Kolmogorov-Smirnov testi ile

c) $N(\mu, \sigma^2)$ dağılımı ile uyumlu olup olmadığına Lilliefors testi ile

d) $N(\mu, \sigma^2)$ dağılımı ile uyumlu olup olmadığına Shapiro-Wilk testi ile

%5 öneme seviyesinde karar veriniz?

Not: 60 40 45 90 70 40 40 70 90 80 70 70 20 10 20 80 60 40 70 90 50
60 80 40 50

SORU:16 İstatistik dersini alan öğrencilerden rastgele seçilen 63 öğrencinin harf notuna göre dağılımı aşağıdadır. Bu örneklem için harf notuna göre öğrencilerin dağılımının kesikli düzgün dağılım gösterip göstermediğine %5 önem seviyesinde karar veriniz?

HARF NOTU	Öğrenci sayısı
AA	5
BA	8
BB	8
CB	7
CC	9
DC	6
DD	9
FD	6
FF	5

SORU:17 A, B, C ve D tedavi yöntemlerinin uygulandığı hastalardan bazıları rastgele seçilmiş ve iyileşme süreleri (gün) aşağıdaki gibi ölçülmüştür. Tedavi yöntemlerine göre iyileşme süresine ait dağılımların homojen varyanslı olup olmadığına %5 önem seviyesinde

a) Brown-Forsythe testini kullanarak

b) Levene testini kullanarak

TEDAVİ YÖNTEMLERİ			
A	B	C	D
10	8	12	14
12	5	10	15
13	6	15	14
14	4	18	12
15		10	18
16		8	
18			

SORU:18 Bir mağaza müşterilerinin %80 den fazlasının kadın olduğunu iddia etmektedir. Mağazaya gelen müşterilerden rastgele seçilen 10 tanesinin cinsiyet dağılımı aşağıdadır. Bu örneklemin Mağaza sahibinin iddiasını destekleyip desteklemediğine %5 önem seviyesinde karar veriniz?

Müşteri no	1	2	3	4	5	6	7	8	9	10
Cinsiyet	E	K	K	K	K	E	E	K	K	E

SORU:19 Örnek hacmi 7 olan bir rastgele örnek için $X_1, X_2, X_3, X_4, X_5, X_6, X_7$ değişkenlerinin sırasıyla 50, 80, 60, 85, 25, 68 ve 86 değerlerini aldığı kabul edelim. Buna göre 1.nci, 5.nci ve 7.nci sıralı istatistiklerinin değerlerini ve 1.nci, 5.inci ve 7.nci sayı sıralı istatistiklerini hesaplayınız?

BÖLÜM 5

İKİ BAĞIMSIZ GRUP (KİTLE) KARŞILAŞTIRMALARINDA KULLANILAN PARAMETRİK OLMAYAN TEST TEKNİKLERİ

Bu bölümde iki bağımsız grubun konum parametrelerinin, dağılım fonksiyonlarının veya değişim parametrelerinin benzerliği yönünden karşılaştırılmasında kullanılan parametrik olmayan tekniklerden bazıları tanıtılacaktır. Bu teknikler aynı amaçla kullanılan *Z testi* ve *t testi* (ortalamaları ve oranları yönünden karşılaştırmada) ve *F testi* (varyansları yönünden karşılaştırmada) olarak bilinen parametrik tekniklerin alternatifleridir. Parametrik tekniklerin uygulanabilmesi için grupların ilgilenilen değişkene göre dağılımlarının normal olması gerektiği varsayımı unutulmamalıdır. Bu varsayımın bir grupta dahi sağlanmaması durumunda parametrik olmayan teknikler tercih edilmelidir.

5.1 KONUM PARAMETRESİ İLE İLGİLİ TESTLER

5.1.1 İKİ BAĞIMSIZ GRUP MEDYAN TESTİ

Bu test iki bağımsız örneğin çekildiği grupların konum parametrelerinin benzerliğini test etmek için kullanılan bir parametrik olmayan test yöntemidir. Konum parametresi olarak medyan parametresini kullanır.

Varsayımları

- i) İlgilenilen değişken sürekli olmalıdır.
- ii) İlgilenilen değişkenin ölçme düzeyi en az sıralama olmalıdır.
- iii) Veriler iki bağımsız gruptan rastgele ve birbirinden bağımsız olarak çekilmelidir.

Test İşleminin Algoritması

1. Hipotezler kurulur

M_1 : Birinci grubun bilinmeyen medyan parametresi

M_2 : İkinci grubun bilinmeyen medyan parametresi olmak üzere;

a) $H_0: M_1 = M_2$

b) $H_0: M_1 = M_2$

c) $H_0: M_1 = M_2$

$H_1: M_1 < M_2$

$H_1: M_1 > M_2$

$H_1: M_1 \neq M_2$

şeklindedir.

2. Gruplardan rastgele ve birbirinden bağımsız olarak örnekler çekilir.

Birinci gruptan n_1 birimlik ve ikinci gruptan da n_2 birimlik örnekler çekilsin. Birinci örnek $X_{11}, X_{12}, \dots, X_{1n_1}$ ve ikinci örnek $X_{21}, X_{22}, \dots, X_{2n_2}$ olsun.

3. İki örnek birleştirilerek birleştirilmiş örnek oluşturulur ve bu örneğin medyan hesaplanır. Birleştirilmiş örnek medyanını m ile gösterelim. Buna göre;

n : Birleştirilmiş örnek hacmi ($n = n_1 + n_2$) olmak üzere, eğer birleştirilmiş örnek birimlerini X_1, X_2, \dots, X_n ile gösterirsek birleştirilmiş örnek medyanı

$$m = \begin{cases} X_{\left(\frac{n+1}{2}\right)}, & n \text{ tek sayı ise} \\ \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2}, & n \text{ çift sayı ise} \end{cases} \quad (5.1)$$

eşitliği ile bulunur.

4. Her iki örnekteki örnek birimleri kendi içlerinde birleştirilmiş örnek medyanından büyük olanlar ile örnek medyanına eşit ve küçük olanlar şeklinde iki alt gruba ayrılır. Bu durumda oluşacak yeni veri düzeni aşağıdaki tabloda gösterildiği gibi olacaktır.

	Örnek No		Toplam
	1	2	
Birleştirilmiş Örnek Medyanından Büyük Değerli Olan Örnek Birimlerinin Sayısı	A	B	A+B
Birleştirilmiş Örnek Medyanına Eşit ve Küçük Değerli Olan Örnek Birimlerinin Sayısı	C	D	C+D
Toplam	A+C= n_1	B+D= n_2	n

H_0 hipotezi doğru iken birinci örnek için $A = C = \frac{n_1}{2}$ ve ikinci örnek için $B = D = \frac{n_2}{2}$ olması beklenir. Aynı zamanda her bir örnek için birleştirilmiş örnek medyanından büyük değerli olanların oranlarının birbirine eşit ya da çok yakın çıkması beklenir.

$P_1 = \frac{A}{n_1}$: Birinci örnekte m 'den büyük değerli olanların oranı ($0 \leq P_1 \leq 1$)

$P_2 = \frac{B}{n_2}$: İkinci örnekte m 'den büyük değerli olanların oranı ($0 \leq P_2 \leq 1$)

5. Test istatistiği belirlenir. Örnek hacimlerinin yeterince büyük olması durumunda Merkezi Limit Teoremi gereğince $P_1 - P_2$ istatistiğinin örnekleme dağılımı normale yaklaşacaktır, öyle ki; $P_1 - P_2 \sim N(E(P_1 - P_2), V(P_1 - P_2))$ dır. Burada H_0 hipotezi doğru iken $E(P_1 - P_2) = 0$ ve $P = \frac{A+B}{n}$ birleştirilmiş örnekte m 'den büyük değerli olanların oranı olmak üzere $V(P_1 - P_2) = P(1 - P) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$ dir. Böylece medyan testi için test istatistiği;

$$Z = \frac{P_1 - P_2 - E(P_1 - P_2)}{\sqrt{V(P_1 - P_2)}} \sim N(0, 1) \quad (5.2)$$

şeklinde tanımlı Z istatistiğidir. H_0 hipotezi doğru iken test istatistiğinin alabileceği değer;

$$Z_h = \frac{P_1 - P_2}{\sqrt{P(1-P) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (5.3)$$

dir.

6. Karar kuralı belirlenerek karar verilir. Karar kuralı α önem seviyesinde H_1 hipotezi dikkate alınarak belirlenir.

Tek yönlü durumda Z_α sağ kritik değer olmak üzere ($P(Z \geq Z_\alpha) = \alpha$ olacak şekilde standart normal dağılım tablosundan Z_α bulunur) $|Z_h| > Z_\alpha$ ise H_0 hipotezi ret edilir, $|Z_h| \leq Z_\alpha$ ise H_0 hipotezi ret edilemez (Veya $P(Z \geq |Z_h|) = p$ olmak üzere $p < \alpha$ ise H_0 hipotezi ret edilir, $p \geq \alpha$ ise H_0 hipotezi ret edilemez).

Çift yönlü durumda $Z_{\frac{\alpha}{2}}$ sağ kritik değer olmak üzere ($P(Z \geq Z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$ olacak şekilde standart normal dağılım tablosundan $Z_{\frac{\alpha}{2}}$ bulunur) $|Z_h| > Z_{\frac{\alpha}{2}}$ ise H_0 hipotezi ret edilir, $-Z_{\frac{\alpha}{2}} \leq Z_h \leq Z_{\frac{\alpha}{2}}$ ise H_0 hipotezi ret edilemez (Veya $P(Z \geq |Z_h|) = p$ olmak üzere $p < \frac{\alpha}{2}$ ise H_0 hipotezi ret edilir, $p \geq \frac{\alpha}{2}$ ise H_0 hipotezi ret edilemez).

SPSS'de Medyan Testi

Spss'de medyan testinin uygulanmasında takip edilecek algoritmanın adımları şu şekilde verilebilir.

1.Adım Variable View sayfasında bağımlı değişken (ilgilenilen değişken) ve grup değişkeni (bağımsız grupları gösteren kategorik değişken) özellikleri ile birlikte tanımlanır. Grup değişkeninin kategorileri **Values** penceresinde belirtilir. **Data View** sayfasında değişkenlere ait veriler girilir. Grup değişkenine ait veriler birinci grup için "1" ve ikinci grup için "2" olarak girilir.

2.Adım Analyze > Nonparametrics Tests > Legacy Dialogs > K Independent Samples... yolu izlenerek **Tests for Several Independent Samples** ekranı açılır. Bu ekranda değişkenler listesinden ilgilenilen değişken seçilerek **Test Variable List** işlem kutusuna, grup değişkeni de **Grouping Variable** işlem kutusuna aktarılır. **Define Groups** penceresinden grup kategorileri girilir.

3.Adım Test Type penceresinden **Median** seçeneği işaretlenir. Tam olasılık değerini elde etmek için **Exact** penceresi tıklanarak açılacak olan pencerede **Exact** seçeneği işaretlenir. **Continue** ve **Ok** tuşları ile test işlemi bitirilir. Sonuçlar çıktı sayfasında tablo halinde sunulur.

4.Adım Çıktı tablosunda test istatistiğinin değeri olarak **Ki-Kare (Chi-Square)** değeri verilir ki bu değer, Z_h^2 değerine eşittir. Ayrıca **Exact. Sig. (2-tailed test)** satırındaki sayısal değer p olasılık değerinin iki katı olup $p = \frac{\text{sayısal değer}}{2}$ hesaplanır. Tek yönlü durumda $p < \alpha$ ise H_0 hipotezi ret edilir, $p \geq \alpha$ ise H_0 hipotezi ret edilemez. Çift yönlü durumda $p < \frac{\alpha}{2}$ ise H_0 hipotezi ret edilir, $p \geq \frac{\alpha}{2}$ ise H_0 hipotezi ret edilemez.

ÖRNEK 5.1 İki farklı tedavi yöntemi ile tedavi edilen hastaların iyileşme sürelerine ait medyanların benzer olup olmadığı araştırılmak istenmektedir. Rastgele seçilen 50 hasta rastgele olarak iki gruba dağıtılmış ve bu gruplara söz konusu tedavi yöntemleri uygulanarak hastaların iyileşme süreleri ölçülmüştür. Gözlenen değerler aşağıdadır. %5 önem seviyesinde tedavi

yöntemlerinin iyileşme süresi üzerine etkilerinin aynı olup olmadığını medyan testi ile kontrol ediniz?

TEDAVİ YÖNTEMLERİ																			
I.YÖNTEM (X_{1i})										II.YÖNTEM (X_{2i})									
18	42	67	18	28	27	11	24	18	16	24	21	7	18	16	40	37	60	70	32
27	18	7	9	41	67	71	18	24	61	30	27	18	29	38	37	40	90	18	71
60	18	24	32	31	61	43	30	18	21										

Cözüm: Bağımlı Değişken (X): İyileşme süresi (gün) ... Nicel, sürekli ve ölçme düzeyi oranlama

Grup Değişkeni (Faktör): Tedavi Yöntemi... Nitel ve ölçme düzeyi sınıflama

Faktör Düzeyleri (kategoriler veya gruplar) $\left\{ \begin{array}{l} I. Grup \dots X_1 (\text{Birinci tedavi yöntemi}) \\ II. Grup \dots X_2 (\text{İkinci Teavi Yöntemi}) \end{array} \right.$

Hipotezler: M_1 - Birinci tedavi yöntemi ile tedavi edilen hastaların iyileşme süresine ait bilinmeyen medyan parametresi

M_2 - İkinci tedavi yöntemi ile tedavi edilen hastaların iyileşme süresine ait bilinmeyen medyan parametresi olmak üzere;

$$H_0: M_1 = M_2$$

$H_1: M_1 \neq M_2$ şeklindedir.

$$\text{Test istatistiği: } Z = \frac{P_1 - P_2 - E(P_1 - P_2)}{\sqrt{V(P_1 - P_2)}} \sim N(0, 1)$$

$n = 50$ çift sayı olup birleştirilmiş örnek medyanı $m = \frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2} = \frac{X_{(25)} + X_{(26)}}{2} = \frac{27+28}{2} = 27,5$ olarak bulunur. Her bir örnek için örnek birimleri $m = 27,5$ dan büyük olanlar ve küçük olanlar şeklinde alt gruplara ayrılırsa aşağıdaki tablo elde edilir.

	Örnek No		Toplam
	1	2	
$m = 27,5$ dan büyük olanların sayısı	A=13	B=12	A+B=25
$m = 27,5$ ve daha küçük olanların sayısı	C=17	D=8	C+D=25
Toplam	A+C= $n_1=30$	B+D= $n_2=20$	$n = 50$

Bu verilere göre: $P_1 = \frac{A}{n_1} = \frac{13}{30} = 0,4333$; $P_2 = \frac{B}{n_2} = \frac{12}{20} = 0,60$ ve $P = \frac{A+B}{n} = \frac{25}{50} = 0,50$ olup, H_0 hipotezi doğru iken $E(P_1 - P_2) = 0$ dir. Ayrıca $V(P_1 - P_2) = P(1 - P) \left(\frac{1}{n_1} + \frac{1}{n_2} \right) = 0,5 * 0,5 \left(\frac{1}{30} + \frac{1}{20} \right) = 0,0208$ ve böylece test istatistiğinin alabileceği değer;

$$Z_h = \frac{0,4333 - 0,60}{\sqrt{0,0208}} = -1,1559 \text{ olarak elde edilir.}$$

Karar $\alpha = 0,05$ ve H_1 hipotezi çift yönlü olduğundan kritik değer $Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$ olup, $|Z_h| = 1,1559 < 1,96$ yani $|Z_h| < Z_{\frac{\alpha}{2}}$ olduğundan H_0 hipotezi ret edilemez. Buna göre tedavi yöntemlerinin hastalarını iyileşme süresi üzerine etkileri arasında istatistiksel olarak anlamlı bir farklılık yoktur.

Spss Çözümü

Hipotezler:

$$H_0: M_1 = M_2$$

$H_1: M_1 \neq M_2$ şeklindedir. Test işlemi sonuç tablosu

$m = 27,5$		tedavi yöntemi	
		yöntem1	yöntem2
iyileşme süresi	> Median	13	12
	<= Median	17	8

Test Statistics^a

		iyileşme süresi	
N		50	$p = \frac{\text{sayısal değer}}{2} = \frac{0,387}{2} = 0,194$ $\alpha = 0,05$ ve H_1 hipotezi çift yönlü olduğundan $\frac{\alpha}{2} = 0,025$ dir. Böylece $p > \frac{\alpha}{2}$ olduğundan H_0 hipotezi ret edilemez. Buna göre tedavi yöntemlerinin hastalarını iyileşme süresi üzerine etkileri arasında istatistiksel olarak anlamlı bir farklılık yoktur.
Median		27,5000	
Chi-Square		1,333	
df		1	
Asymp. Sig.		,248	
Yates' Continuity Correction	Chi-Square	,750	
	df	1	
Asymp. Sig.		,386	
Exact Sig.		,387	
Point Probability		,239	

a. Grouping Variable: tedavi yöntemi