T.C.

ONDOKUZ MAYIS ÜNİVERSİTESİ

FEN EDEBİYAT FAKÜLTESİ
İSTATİSTİK BÖLÜMÜ

İST.482 PARAMETRİK OLMAYAN İSTATİSTİKSEL YÖNTEMLER

PROF. DR. YÜKSEL ÖNER

8. Hafta

5.1.4 KOLMOGOROV-SMIRNOV TESTİ

İki bağımsız grubu dağılım fonksiyonlarının benzerliği yönünden karşılaştırmak amacı ile kullanılan bir diğer parametrik olmayan tekniktir.

Varsayımları

i) İlgilenilen değişken süreklidir

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

$F_1(x)$: Birinci gruba ait bilinmeyen dağılım fonksiyonu

$F_2(x)$: İkinci gruba ait bilinmeyen dağılım fonksiyonu olmak üzere

a) H_0: $F_1(x) = F_2(x)$, bütün x değerleri için

H_1: $F_1(x) < F_2(x)$, en az bir x değerleri için

b) H_0: $F_1(x) = F_2(x)$, bütün x değerleri için

H_1: $F_1(x) > F_2(x)$, en az bir x değerleri için

c) H_0: $F_1(x) = F_2(x)$, bütün x değerleri için

H_1: $F_1(x) \neq F_2(x)$, en az bir x değerleri için

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 örnekte örnek birimleri küçükten büyüğe doğru sıralanır ve her bir örnek için ayrı ayrı örnek dağılım fonksiyonları bulunur.

Birinci örneğin dağılım fonksiyonu $S_1(x) = \dfrac{\text{Birinci örnekte } x \text{ değerine eşit veya daha küçük değerli gözlemlerin sayısı}}{n_1}$

İkinci örneğin dağılım fonksiyonu $S_2(x) = \dfrac{\text{İkinci örnekte } x \text{ değerine eşit veya daha küçük değerli gözlemlerin sayısı}}{n_2}$

4. Test istatistiği belirlenir. İki örnek Kolmogorov-Smirnov testi için test istatistiği;

$$D = Enb|S_1(x) - S_2(x)| \qquad (5.6)$$

şeklinde tanımlıdır. Test istatistiğinin örnekten hesaplanan değeri D_h olsun.

5. Karar kuralı belirlenir ve karar verilerek değerlendirilir. H_0 hipotezi doğru iken test istatistiğinin sıfır veya yeteri kadar küçük değer alması beklenir. H_1 hipotezine göre ise test istatistiğinin büyük değer alması beklenir. Söz konusu olan büyüklük kriteri test istatistiğinin örnekleme dağılımından H_1 hipotezine göre belirlenecek olan D_k kritik değeridir. D_k kritik değeri $n_1 = n_2 = n,\ H_1$ tek veya çift yanlı ve $1 - \alpha$ için (T13(a)) tablosundan belirlenirken, $n_1 \neq n_2,\ H_1$ tek veya çift yanlı ve $1 - \alpha$ için (T13(b)) tablosundan belirlenmektedir. Eğer $D_h \geq D_k$ ise H_0 hipotezi ret edilir, $D_h < D_k$ ise H_0 hipotezi ret edilemez.

SPSS'de İki Örnek Kolmogorov-Smirnov Testi

Spss'de iki örnek Kolmogorov-Smirnov 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 > 2 Independent Samples… yolu izlenerek **Two Independent Samples Tests** 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 **Kolmogorov-Smirnov Z** 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 **Most Extreme Absolute** alınır. **Exact Sig.(2 tailed)** satırında olasılık değeri p olarak alınırsa, H_1 hipotezi tek yönlü iken $\frac{p}{2} > \alpha$ ise H_0 hipotezi ret edilemez, $\frac{p}{2} \leq \alpha$ ise H_0 hipotezi ret edilir. H_1 hipotezi çift yönlü iken $p > \alpha$ ise H_0 hipotezi ret edilemez, $p \leq \alpha$ ise H_0 hipotezi ret edilir.

Örnek 5.8 Bir eğitim programı ile belirli dönemlerde çok sayıda öğrenci sınava hazırlanmaktadır. Eğitim programını tamamlayan kız öğrencilerden 6'sı ve erkek öğrencilerden de 7'si rastgele seçilerek ilgili konuda hazırlanan bir teste tabi tutulmuşlardır. Test sonucunda alınan puanlar cinsiyete göre aşağıdadır. Kız ve erkek öğrencilerin başarı dağılımlarının benzer olup olmadığına %10 önem seviyesinde karar veriniz.

Kız(X_{1i})	40 50 54 60 70 84
Erkek(X_{2i})	45 56 59 62 64 75 90

Çözüm: Bağımlı değişken: Başarı Puanı…. Nicel, sürekli ve ölçme düzeyi eşit aralıklı

Faktör (Gruplama değişkeni): Cinsiyet… Nitel ve ölçme düzeyi sınıflama

I. Grup… Kız öğrenciler

II. Grup… Erkek öğrenciler olup, gruplar bağımsızdır.

Hipotezler kurulur.

$F_1(x)$: Kızlar grubuna ait başarı notları için bilinmeyen dağılım fonksiyonu

$F_2(x)$: Erkekler grubuna ait başarı notları için bilinmeyen dağılım fonksiyonu olmak üzere

$H_0: F_1(x) = F_2(x)$, bütün x değerleri için

$H_1: F_1(x) \neq F_2(x)$, en az bir x değerleri için

şeklindedir. İki örneklemi birleştirerek birleştirilmiş örnekte örnek birimlerini küçükten büyüğe doğru sıralayalım ve her bir örnek için örnek dağılım fonksiyonunu oluşturalım. Test istatistiği;

$D = Enb|S_1(x) - S_2(x)|$ olup, alabileceği değeri bulalım.

X_{1i}	X_{2i}	$S_1(x_{1i})$	$S_2(x_{2i})$	$\lvert S_1(x) - S_2(x) \rvert$
40		1/6	0	7/42
	45	1/6	1/7	1/42
50		2/6	1/7	8/42
54		3/6	1/7	15/42
	56	3/6	2/7	9/42
	59	3/6	3/7	3/42
60		4/6	3/7	10/42
	62	4/6	4/7	4/42
	64	4/6	5/7	2/42
70		5/6	5/7	5/42
	75	5/6	6/7	1/42
84		6/6	6/7	6/42
	90	6/6	7/7	0,00

$D_h = 15/42$, $n_1 = 6$ ve $n_2 = 7$, $1 - \alpha = 0,90$ ve H_1 çift yönlü olduğundan T13(b)'den kritik değer $D_k = \frac{4}{7}$ =24/42 olup, $D_h < D_k$ dır. Bu sebeple H_0 hipotezi ret edilemez. Buna göre Kız ve erkek öğrencilerin başarı notlarına ait dağılımlar benzerdir.

Spss Çözümü Hipotezler

$H_0: F_1(x) = F_2(x)$, bütün x değerleri için

$H_1: F_1(x) \neq F_2(x)$, en az bir x değerleri için

Test Statistics[a]

		puan
Most Extreme Differences	Absolute	,357
	Positive	,048
	Negative	-,357
Kolmogorov-Smirnov Z		,642
Asymp. Sig. (2-tailed)		,804
Exact Sig. (2-tailed)		,712
Point Probability		,089

$D_h = 0,357 = \frac{15}{42}$, $\alpha = 0,10$ ve $p = 0,712$ olup, $p > \alpha$ olduğundan sebeple H_0 hipotezi ret edilemez. Buna göre Kız ve erkek öğrencilerin başarı notlarına ait dağılımlar benzerdir.

Örnek 5.9 Bir hastalığın tedavisinde *A* ve *B* gibi iki farklı yöntem uygulanmaktadır. Uzmanlar *B* yönteminin uygulanması durumunda hastaların daha kısa sürede iyileşme göstereceklerini iddia etmektedirler. Bu durumun geçerliliğini kontrol etmek amacı ile hastalık dereceleri aynı olan 14 hasta rastgele seçilmiş ve her birinde 7 hasta olacak şekilde hastalar rastgele olarak iki gruba ayrılmıştır. Birinci gruba *A* tedavi yöntemi ve ikinci gruba *B* tedavi yöntemi uygulanarak tedavi edilen hastaların iyileşme süreleri (gün) aşağıdaki gibi gözlenmiştir. Bu verilere göre iddianın geçerliliğini %5 önem seviyesinde:

a) Mann Whitney U testi ile **b)** Wald –Wolfowitz dizi parçaları testi ile

c) Kolmogorov- Smirnov testi ile karar veriniz?

***A* Yöntemi**	25	24	21	22	23	19	20
$\boldsymbol{R(X_{1i})}$	14	13	10	11	12	8	9
***B* Yöntemi**	11	10	15	12	18	14	13
$\boldsymbol{R(X_{2i})}$	2	1	6	3	7	5	4

Çözüm: Değişken (X): İyileşme süresi(gün)…. Nicel türden, sürekli ve ölçme düzeyi oranlama

Grup değişkeni (Faktör); Tedavi yöntemi… Nitel türden ve ölçme düzeyi sınıflama

$$\left.\begin{array}{l} I.\ Grıp: A\ tedavi\ yöntemi\ (1) \\ II. Grup: B\ tedavi\ yöntemi\ (2) \end{array}\right\}\text{Bağımsız gruplar}$$

a) Mann Whitney U testi iki bağımsız grubu medyanları yönünden karşılaştırır.

M_1: A Yöntemi (I.grup) için medyan parametresi

M_2: B Yöntemi (II.grup) için medyan parametresi olmak üzere hipotezler:

$H_0: M_1 = M_2$

$H_1: M_1 > M_2$

Test istatistiği: $T = S - \frac{n_1(n_1+1)}{2}$ ve $S = \sum_{i=1}^{n_1} R(X_{1i}) = 77$ ve böylece test istatistiğinin alabileceği değer $T_h = 77 - \frac{7*8}{2} = 49$ bulunur.

Karar: $\alpha = 0{,}05$ önem seviyesinde H_1 hipotezine göre karar kuralı $T_h \geq W'_\alpha$ ise H_0 hipotezi ret edilir, aksi takdirde ret edilemez. Kritik değer, $W'_\alpha = n_1 * n_2 - W_\alpha$ ve $n_1 = n_2 = 7$ ve $\alpha = 0{,}05$ için $W_\alpha = 12$, $W'_\alpha = 49 - 12 = 37$ olur. Böylece 49>37, yani $T_h \geq W'_\alpha$ olduğundan H_0 hipotezi ret edilir. Bu durumda B tedavi yöntemi ile tedavi edilen hastaların iyileşme süreleri A yöntemi ile tedavi edilenlere göre daha kısa sürelidir.

Spss çözümü:

	tedaviyöntem	N	Mean Rank	Sum of Ranks
süre	A Yöntemi	7	11,00	77,00
	B Yöntemi	7	4,00	28,00
	Total	14		

	süre	
Mann-Whitney U	,000	$T_h = 0$ olup, elle çözünden elde edilen $T_h = 49$ değerinden farklıdır. Bunun nedeni Spss bu değeri hesaplarken Sdeğeri küçük olan grubu dikkate almasıdır. Yani $S = 28$ olup .$T_h = S - \frac{n_2(n_2+1)}{2} = 28 - \frac{7*8}{2} = 0$ bulunur.H_1 tek yönlü olduğundan $p = 0{,}001$ dir. $\alpha = 0{,}05$ önem seviyesinde iken $p < \alpha$ olduğundan H_0 ret edilir. Sonuç olarak B tedavi yöntemi için iyileşme süresi A tedavi yöntemine göre daha kısadır.
Wilcoxon W	28,000	
Z	-3,130	
Asymp. Sig. (2-tailed)	,002	
Exact Sig. [2*(1-tailed Sig.)]	,001	

b) Wald-Wolfowitz dizi parçaları testi iki bağımsız grubu dağılım fonksiyonları yönünden karşılaştırır.
$F_1(x)$: A tedavi yöntemi için iyileşme süresine ait bilinmeyen dağılım fonksiyonu

$F_2(x)$: B tedavi yöntemi için iyileşme süresine ait bilinmeyen dağılım fonksiyonu olmak üzere, hipotezler

H_0: $F_1(x) = F_2(x)$, bütün x değerleri için

H_1: $F_1(x) > F_2(x)$, en az bir x değerleri için

İki örnek birleştirilerek, birleştirilmiş örnekte örnek birimleri küçükten büyüğe doğru sıralanır ve her bir örnek birimi ait olduğu gruba göre, A tedavi yönteminin kullanıldığı gruba aitse (+) ve B tedavi yönteminin kullanıldığı gruba aitse (-) ile simgelenir.

X_i :	10	11	12	13	14	15	18	19	20	21	22	23	24	25
Simge :	-	-	-	-	-	-	-	+	+	+	+	+	+	+

Test istatistiği ve alabileceği değer: r: Dizi parçalarının sayısı olup, $r_h = 2$ dır.
Karar: $\alpha = 0{,}05$ ve $n_1 = n_2 = 7$ iken kritik değeri (T11.a) dan $r_{\alpha/2} = 3$ olup, $r_h = 2 < 3 = r_{\alpha/2}$ olduğundan H_0 hipotezi ret edilir. Bu durumda B tedavi yöntemi ile tedavi edilen hastaların iyileşme süreleri A yöntemi ile tedavi edilenlere göre daha kısa sürelidir.

Test Statistics

		Number of Runs	Z	Exact Sig. (1-tailed)	$\alpha = 0{,}025$ önem seviyesinde p =,001 < α olduğundan H_0 hipotezi ret edilir. Sonuç olarak B tedavi yöntemi için iyileşme süresi A tedavi yöntemine göre daha kısadır.
süre	Exact Number of Runs	r_h =2	-3,060	,001	

c) Kolmogorov Smirnov testi iki bağımsız grubu dağılım fonksiyonları yönünden karşılaştırır.
$F_1(x)$: A tedavi yöntemi için iyileşme süresine ait bilinmeyen dağılım fonksiyonu

$F_2(x)$: B tedavi yöntemi için iyileşme süresine ait bilinmeyen dağılım fonksiyonu olmak üzere, hipotezler

H_0: $F_1(x) = F_2(x)$, bütün x değerleri için
H_1: $F_1(x) > F_2(x)$, en az bir x değerleri için
şeklindedir. İki örneklemi birleştirerek birleştirilmiş örnekte örnek birimlerini küçükten büyüğe doğru sıralayalım ve her bir örnek için örnek dağılım fonksiyonunu oluşturalım. Test istatistiği;

$D = Enb|S_1(x) - S_2(x)|$ olup, alabileceği değeri bulalım.

X_{1i}	X_{2i}	$S_1(x_{1i})$	$S_2(x_{2i})$	$\|S_1(x) - S_2(x)\|$
	10	0	1/7	1/7
	11	0	2/7	2/7
	12	0	3/7	3/7
	13	0	4/7	4/7
	14	0	5/7	5/7
	15	0	6/7	6/7
	18	0	1	1
19		1/7	1	6/7
20		2/7	1	5/7
21		3/7	1	4/7
22		4/7	1	3/7
23		5/7	1	2/7
24		6/7	1	1/7
25		1	1	0

$D_h = 1$, $n_1 = 7$ ve $n_2 = 7$, $1 - \alpha = 0{,}95$ ve H_1 tek yönlü olduğundan T13(a) dan kritik değer $D_k = \frac{4}{7}$ olup, $D_h > D_k$ dır. Bu sebeple H_0 hipotezi ret edilir. Bu durumda B tedavi yöntemi ile tedavi edilen hastaların iyileşme süreleri A yöntemi ile tedavi edilenlere göre daha kısa sürelidir.

Test Statistics

		süre
Most Extreme Differences	Absolute	1,000
	Positive	,000
	Negative	-1,000
Kolmogorov-Smirnov Z		1,871
Asymp. Sig. (2-tailed)		,002

$D_h = 1{,}000$, $\alpha = 0{,}05$ ve $p = 0{,}002$ olup, $p < \alpha$ olduğundan sebeple H_0 hipotezi ret edilir. Buna göre B tedavi yöntemi ile tedavi edilen hastaların iyileşme süreleri A yöntemi ile tedavi edilenlere göre daha kısa sürelidir.

5.1.5 FISHER TAM OLASILIK TESTİ (FISHER EXACT TEST)

İki bağımsız grubu iki değer alan bir değişken yönünden ilgilenilen bir özelliğe sahip olanların oranları bakımından karşılaştırmak amacı ile kullanılan testlerden birisidir. İki bağımsız gruba ait verilerde her bir birimin ilgilenilen özelliğe sahip olma durumuna göre iki ayrık sınıftan birisine sınıflandırılması sık karşılaşılan uygulamalardandır. Örneğin; A ve B ilaçlarının uygulandığı hastaların tedaviye olumlu tepki verenler ve vermeyenler olarak sınıflandırılması, A ve B türü tohumların çimlenme durumuna göre çimlenenler ve çimlenmeyenler olarak sınıflandırılması v.s. düşünülebilir. Bu tür sınıflandırmalarda veri düzeni 2*2 çapraz frekans tablosu veri düzeni olup aşağıdaki tablo ile verilir.

Örnek No	İlgilenilen özellikte olanlar	İlgilenilen özellikte olmayanlar	Toplam
1	a	$n_1 - a$	n_1
2	x	$n_2 - x$	n_2
Toplam	a+x	n −(a + x)	n

Burada amaç iki ayrık sınıfa düşen birimlerin oranı bakımından iki grubun farklılık gösterip göstermediğini belirlemektir. Bu amaçla uygulanacak olan hipotez testinde test edilecek **Hipotez grubu**:

Π_1: Birinci grupta ilgilenilen özelliğe sahip olanların oran parametresi
Π_2: İkinci grupta ilgilenilen özelliğe sahip olanların oran parametresi olmak üzere

a) $H_0: \Pi_1 = \Pi_2$
$H_1: \Pi_1 > \Pi_2$

b) $H_0: \Pi_1 = \Pi_2$
$H_1: \Pi_1 < \Pi_2$

c) $H_0: \Pi_1 = \Pi_2$
$H_1: \Pi_1 \neq \Pi_2$

şeklindedir.

Test istatistiği; ikinci örnek üzerinde tanımlanan

$$b: \text{İ}kinci\ \ddot{o}rnekte\ ilgilenilen\ \ddot{o}zellikte\ olan\ birimlerin\ say\imath s\imath \qquad (5.7)$$

istatistiğidir. Bu istatistiğin alabileceği değerler $x = 0, 1, 2, \ldots, n_2$ olup, b istatistiğinin dağılımı hipergeometrik dağılım gösterir ve olasılık fonksiyonu;

$$f(x, a) = P(b = x) = \frac{\binom{n_2}{x}\binom{n_1}{a}}{\binom{n_1+n_2}{a+x}}, x = 0, 1, 2, \ldots, n_2; a + x > 0 \qquad (5.8)$$

şeklindedir. Test istatistiğinin örnekten hesaplanan değeri b_h olsun.

Karar: α önem seviyesinde H_1 hipotezine göre karar kuralı belirlenir.

H_1	**Karar kuralı**
$\Pi_1 > \Pi_2$	$b_h \leq b_\alpha$ ise H_0 hipotezi ret edilir…..$P(b \leq b_\alpha) = \alpha$ $b_h > b_\alpha$ ise H_0 hipotezi ret edilemez
$\Pi_1 < \Pi_2$	$b_h \geq b'_\alpha$ ise H_0 hipotezi ret edilir…..$P(b \geq b'_\alpha) = \alpha$ $b_h < b'_\alpha$ ise H_0 hipotezi ret edilemez
$\Pi_1 \neq \Pi_2$	$b_h \leq b_{\alpha/2}$ veya $b_h \geq b'_{\alpha/2}$ ise H_0 hipotezi ret edilir $b_{\alpha/2} < b_h < b'_{\alpha/2}$ ise H_0 hipotezi ret edilemez…. $P(b \leq b_{\alpha/2}) = \alpha/2$ ve …..$P(b \geq b'_{\alpha/2}) = \alpha/2$ dir.

SPSS'de Fisher Tam (Kesin) Olasılık Testi

1.Adım Variable View sayfasında değişken olarak örnek, özellik, frekans olmak üzere üç değişken ve bu değişkenlerin özellikleri tanımlanır. Örnek kategorik bir değişken olup kategorileri birinci örnek için (1) ikinci örnek için (2) olarak **Values** bölümünde kodlanır. Özellik ayrık iki değer alan bir kategorik değişkendir ve kategorileri ilgilenilen özelliğe sahip olan birimler için (1) ilgilenilen özelliğe sahip olmayan birimler için (2) olarak **Values** bölümünde kodlanır. Frekans değişkeni nicel bir değişkendir. **Data View** sayfasında değişkenlere ait veriler girilir. Örnek değişkenine ait veriler birinci örnek için "1" ve ikinci örnek için "2" olarak, Özellik değişkenine ait veriler ilgilenilen özellik için "1" ve ilgilenilmeyen özellik için "2" olarak ve frekans değişkeni için çapraz tablonun her bir hücresindeki frekans değerleri ilgili yerlere girilir.

2.Adım Frekans değişkeni için ağırlıklandırma işlemi uygulanır.

3. Adım Analyze > Descriptive Statistics > Crosstabs… yolu izlenerek **Crosstabs** ekranı açılır. Bu ekranda değişkenler listesinden örnek değişkeni seçilerek **Row**(satır) işlem kutusuna, özellik değişkeni seçilerek **Column** (sütun) işlem kutusuna aktarılır. Bu ekranda bulunan **Statistics** penceresi açılarak **Chi-square** seçeneği işaretlenir. **Continue** ve **Ok** tuşları ile test işlemi bitirilir. Sonuçlar çıktı sayfasında tablo halinde sunulur.

4.Adım Çıktı tablolarında verilere ait çapraz tablo ve test işlem sonuç tablosu yer almaktadır. Test işlemi sonuç tablosunda Fisher tam olasılık testi için p olasılığı tek yönlü testlerde **Exact Sig.(1 sided)** sütunundan, çift yönlü testlerde ise **Exact Sig.(2 sided)** sütunundan okunur. Eğer $p > \alpha$ ise H_0 hipotezi ret edilemez, $p \leq \alpha$ ise H_0 hipotezi ret edilir.

Örnek 5.10 Aşağıdaki 2*2 veri düzenini kullanarak "b: İkinci örnekte ilgilenilen özellikte olan birimlerin sayısı " olarak tanımlansın.

a) b istatistiğinin örnekleme dağılımını oluşturunuz?

b) $\boldsymbol{b}$ istatistiğinin beklenen değer ve varyansını hesaplayınız?

c) %4 önem seviyesinde H_1: $\Pi_1 > \Pi_2$ iken, H_1: $\Pi_1 < \Pi_2$ iken ve H_1: $\Pi_1 \neq \Pi_2$ iken kritik değer/değerler ne olmalıdır?

Örnek No	İlgilenilen Özellikte Olanlar	İlgilenilen özellikte Olmayanlar	Toplam
1	$a = 9$	$n_1 - a$=1	n_1=10
2	$x = 1$	$n_2 - x$=3	n_2=4
Toplam	$a + x = 10$	$n - (a + x)$	n=14

Çözüm: a) Veriler iki örnekten oluşmakta ve her bir örnekte örnek birimleri iki ayrık sınıftan birine sınıflanmaktadır. Test istatistiği;

b : İkinci örnekte ilgilenilen özellikte olan birimlerin sayısı

olmak üzere, örnekleme dağılımı hipergeometrik dağılımdır. Çünkü n tane birim, birinden n_1 tane diğerinden n_2 tane olmak üzere iki tür nesneden oluşmakta ve n tane nesne içerisinden rastgele çekilen n_2 tane nesnenin içerisinde ilgilenilen özelliği sahip olanların sayısının b olması durumu ve bu olayın gerçekleşmesi olasılığı ile ilgileniyoruz. Buna göre b istatistiğinin olasılık fonksiyonu;

$$f(x, a) = P(b = x) = \frac{\binom{n_2}{x}\binom{n_1}{a}}{\binom{n_1+n_2}{a+x}} \text{ , } x = 0, 1, 2, \ldots, n_2 \text{ve } a + x = 10, n_2 = 4$$

$$= \frac{\binom{4}{x}\binom{10}{a}}{\binom{14}{a+x}} \text{ ; } x = 0,1,2,3,4 \text{ ve } a + x = 10$$

olur. Bu durumda b istatistiğinin alabileceği değerler ve bu değerleri alma olasılıkları aşağıda verilmiştir.

$b=x$	$P(b=x)$	$b=x$	$P(b=x)$
0	$P(b=0)=\frac{\binom{4}{0}\binom{10}{10}}{\binom{14}{10}}=\frac{1}{1001}$	3	$P(b=3)=\frac{\binom{4}{3}\binom{10}{7}}{\binom{14}{10}}=\frac{480}{1001}$
1	$P(b=1)=\frac{\binom{4}{1}\binom{10}{9}}{\binom{14}{10}}=\frac{40}{1001}$	4	$P(b=4)=\frac{\binom{4}{4}\binom{10}{6}}{\binom{14}{10}}=\frac{210}{1001}$
2	$P(b=2)=\frac{\binom{4}{2}\binom{10}{8}}{\binom{14}{10}}=\frac{270}{1001}$		

b) b istatistiğinin beklenen değeri: $E(b)=\sum_{x=0}^{4} xP(b=x)=\frac{1}{1001}(0*1+1*40+2*270+3*480+4*210)=2{,}857$

b istatistiğinin varyansı ise $V(b)=E(b^2)-[E(b)]^2=\sum_{x=0}^{4} x^2P(b=x)-((2{,}857)^2)$

$=\frac{1}{1001}(0^2*1+1^2*40+2^2*270+3^2*480+4^2*210)-((2{,}857)^2)=0{,}629$ bulunur.

c) $\boldsymbol{H_1}$: $\boldsymbol{\Pi_1>\Pi_2}$ **iken** $P(b\le b_\alpha)=\alpha=0{,}04$ ise $b_\alpha=?$

$b_\alpha=0$ ise $P(b\le b_\alpha)=P(b=0)=\frac{1}{1001}=0{,}001<\alpha$

$b_\alpha=1$ ise $P(b\le b_\alpha)=P(b=0)+P(b=1)=\frac{1}{1001}+\frac{40}{1001}=\frac{41}{1001}=\mathbf{0{,}041}>\alpha$ $\Rightarrow 0{,}041$ değeri $\alpha=0{,}04$ değerine daha yakın olduğundan $b_\alpha=1$ olmalıdır.

$\boldsymbol{H_1}$: $\boldsymbol{\Pi_1<\Pi_2}$ **iken** $P(b\ge b'_\alpha)=\alpha=0{,}04$ ise $b'_\alpha=?$

$b'_\alpha=4$ ise $P(b\ge b'_\alpha)=P(b\ge 4)=\frac{210}{1001}=\mathbf{0{,}209}>\alpha$ $\Rightarrow$ b istatistiğinin alabileceği en büyük değer 4 olduğu için $b'_\alpha=4$ olacaktır.

$\boldsymbol{H_1}$: $\boldsymbol{\Pi_1\ne\Pi_2}$ **iken** $P\left(b\le b_{\alpha/2}\right)=\alpha/2=0{,}02$ ise $b_{\alpha/2}=?$ ve $P\left(b\ge b'_{\alpha/2}\right)={}^{\alpha}/_{2}=0{,}02$ ise $b'_{\alpha/2}=?$

$b_{\alpha/2}=0$ ise $P\left(b\le b_{\frac{\alpha}{2}}\right)=P(b=0)=\frac{1}{1001}=\mathbf{0{,}001}<\frac{\boldsymbol{\alpha}}{\mathbf{2}}=\mathbf{0{,}02}$

$b_{\alpha/2}=1$ ise $P\left(b\le b_{\frac{\alpha}{2}}\right)=P(b\le 1)=P(b=0)+P(b=1)=\frac{1}{1001}+\frac{40}{1001}=\frac{41}{1001}=0{,}041>\frac{\alpha}{2}=0{,}02$ $\Rightarrow 0{,}001$ değeri $\frac{\alpha}{2}=0{,}02$ değerine daha yakın olduğundan $b_{\frac{\alpha}{2}}=0$ olmalıdır.

$b'_{\frac{\alpha}{2}}=4$ ise $P\left(b\ge b'_{\frac{\alpha}{2}}\right)=P(b\ge 4)=\frac{210}{1001}=\mathbf{0{,}209}>\frac{\boldsymbol{\alpha}}{\mathbf{2}}=\mathbf{0{,}02}$ $\Rightarrow$ b istatistiğinin alabileceği en büyük değer 4 olduğu için $b'_{\frac{\alpha}{2}}=4$ olacaktır.

Örnek 5.11 A ve B türü tohumlardan rastgele seçilen 12'şer tohumun çimlenme durumuna göre dağılımı aşağıda verilmiştir. Bu bilgiye göre %3 önem seviyesinde A ve B türü tohumlarda çimlenmeyenlerin oranlarının farklı oldukları söylenebilir mi?

	Çimlenmedi	**Çimlendi**	**Toplam**
Tohum A	a=5	$n_1 - a$ =7	n_1 =12
Tohum B	x=3	$n_2 - x$ =9	n_2 =12
Toplam	$a + x$=8	$n -(a\ +\ x)$=16	n =24

Çözüm: Değişken(X): Tohumun çimlenme durumu…. Nitel ve ölçme düzeyi sınıflama

İlgilenilen özellik: Tohumun çimlenmemesi (1)

İlgilenilmeyen özellik: Tohumun çimlenmesi (2)

Birinci Grup: A türü tohum ve İkinci grup: B türü tohum olup gruplar birbirinden bağımsızdır.

Π_1: A türü tohumda çimlenmeyenlerin oranı

Π_2: B türü tohumda çimlenmeyenlerin oranı, olmak üzere hipotezler:

$H_0 : \Pi_1 = \Pi_2$

$H_1 : \Pi_1 \neq \Pi_2$ şeklinde oluşturulur.

Test istatistiği; b : İkinci örnekte (yani B tohum türünde) çimlenmeyenlerin sayısı

olmak üzere, test istatistiğinin örnekten hesaplanan değeri $b_h = 3$ dür.

Karar: $\alpha = 0{,}03$ önem seviyesinde H_1'e göre karar kuralı; $b_h \leq b_{\alpha/2}$ veya $b_h \geq b'_{\alpha/2}$ ise H_0 hipotezi ret edilir ve $b_{\alpha/2} < b_h < b'_{\alpha/2}$ ise H_0 hipotezi ret edilemez. Burada $P\left(b \leq b_{\alpha/2}\right) = \alpha/2$ ve $P\left(b \geq b'_{\alpha/2}\right) = {}^{\alpha}/_{2}$ olacak şekilde $b_{\alpha/2}$ ve $b'_{\alpha/2}$ kritik değerleri bulunmalıdır.

b istatistiğinin örnekleme dağılımı için olasılık fonksiyonu;

$$f(x,a) = P(b = x) = \frac{\binom{n_2}{x}\binom{n_1}{a}}{\binom{n_1+n_2}{a+x}} \quad , x = 0, 1, 2, \ldots, n_2 \text{ ve a=5, } a + x = 8, n_2 = 12$$

$$= \frac{\binom{12}{x}\binom{12}{a}}{\binom{24}{8}} \; ; x = 0,1,2,3, \ldots, 12 \text{ ve } a + x = 8$$

dir. Buna göre; $\alpha = 0{,}03$ ve $\frac{\alpha}{2} = 0{,}015$

$$P(b = 0) = \frac{\binom{12}{0}\binom{12}{8}}{\binom{24}{8}} = \frac{495}{735471} = 0{,}00067 \quad ; \quad P(b = 1) = \frac{\binom{12}{1}\binom{12}{7}}{\binom{24}{8}} = \frac{9504}{735471} = 0{,}0129$$

$$P(b = 2) = \frac{\binom{12}{2}\binom{12}{6}}{\binom{24}{8}} = \frac{60984}{735471} = 0{,}083 \quad ; \quad P(b = 6) = \frac{\binom{12}{6}\binom{12}{2}}{\binom{24}{8}} = \frac{60984}{735471} = 0{,}083$$

$$P(b = 7) = \frac{\binom{12}{7}\binom{12}{1}}{\binom{24}{8}} = \frac{9504}{735471} = 0{,}0129 \quad ; \quad P(b = 8) = \frac{\binom{12}{8}\binom{12}{0}}{\binom{24}{8}} = \frac{495}{735471} = 0{,}00067$$

bulunur. Böylece;

$P(b \le 1) = P(b=0) + P(b=1) = 0{,}00067 + 0{,}0129 = \mathbf{0{,}0137} < \frac{\alpha}{2}$ **=0,015…(*)**

$P(b \le 2) = 0{,}0137 + P(b=2) = 0{,}0137 + 0{,}083 = 0{,}0967 > \frac{\alpha}{2} = 0{,}015$

(*) eşitsizliğine göre $b_{\alpha/2} = 1$ bulunur. Benzer şekilde;

$P(b \ge 7) = P(b=7) + P(b=8) = 0{,}0129 + 0{,}00067 = \mathbf{0{,}0137} < \frac{\alpha}{2}$ **=0,015…(**)**

$P(b \ge 6) = P(b=6) + 0{,}0137 = 0{,}083 + 0{,}0137 = 0{,}0967 > \frac{\alpha}{2} = 0{,}015$

(**) eşitsizliğine göre $b'_{\alpha/2} = 7$ bulunur. Sonuç olarak $1 < 3 < 7$ yani$b_{\alpha/2} < b_h < b'_{\alpha/2}$ olduğundan H_0 hipotezi ret edilemez. O halde A ve B tohum türlerinde çimlenmeme oranları benzerdir.

Spss çözümü: Değişken(X): Tohumun çimlenme durumu…. Nitel ve ölçme düzeyi sınıflama

İlgilenilen özellik: Tohumun çimlenmemesi (1)

İlgilenilmeyen özellik: Tohumun çimlenmesi (2)

Birinci Grup: A türü tohum ve İkinci grup: B türü tohum olup gruplar birbirinden bağımsızdır.

Π_1: A türü tohumda çimlenmeyenlerin oranı

Π_2: B türü tohumda çimlenmeyenlerin oranı, olmak üzere hipotezler:

$H_0 : \Pi_1 = \Pi_2$

$H_1 : \Pi_1 \neq \Pi_2$ şeklinde oluşturulur.

Test istatistiğinin örnekten hesaplanan değeri $b_h = 3$ ve H_1 çift yönlü olduğundan $p = 0{,}667$ olup, ; $\alpha = 0{,}03$ iken $p > \alpha$ dır. Bu sebeple H_0 hipotezi ret edilemez. O halde A ve B tohum türlerinde çimlenmeme oranları benzerdir.

tohum * durum Crosstabulation

		durum		Total
		çimlenmedi	çimlendi	
tohum	A tohumu	5	7	12
	B tohumu	3	9	12
Total		8	16	24

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,750[a]	1	,386		
Continuity Correction	,188	1	,665		
Likelihood Ratio	,756	1	,385		
Fisher's Exact Test				,667	,333
Linear-by-Linear Association	,719	1	,397		
N of Valid Cases	24				

<u>Örnek 5.12</u> İlaçla tedavi edilen hastalardan 8 ve cerrahi tedavi uygulanan hastalardan 8 tanesi rastgele seçilmiş ve bunlar tedaviye tepki verme durumlarına göre aşağıdaki gibi bir dağılım göstermişlerdir. %6 önem seviyesinde tedaviye olumsuz tepki verenlerin oranının cerrahi tedavi uygulanan hastalarda daha düşük olduğu söylenebilir mi?

	Olumsuz	**Olumlu**	**Toplam**
İlaçla Tedavi	a=5	$n_1 - a$ =3	n_1 =8
Cerrahi Tedavi	x=1	$n_2 - x$ =7	n_2 =8
Toplam	$a + x$=6	$n - (a + x)$=10	n =16

<u>Çözüm:</u> **<u>:</u>** Değişken(X): Tedaviye tepki verme durumu…. Nitel ve ölçme düzeyi sınıflama

İlgilenilen özellik: Olumsuz tepki verme (1)

İlgilenilmeyen özellik: Olumlu tepki verme (2)

Birinci Grup: İlaçla tedavi yöntemi ve İkinci grup: Cerrahi tedavi yöntemi olup gruplar birbirinden bağımsızdır.

Π_1: İlaçla tedavi yönteminde olumsuz tepki verenlerin oranı

Π_2: Cerrahi yönteminde olumsuz tepki verenlerin oranı, olmak üzere hipotezler:

$H_0 : \Pi_1 = \Pi_2$

$H_1 : \Pi_1 > \Pi_2$ şeklinde oluşturulur.

Test istatistiği; b : İkinci örnekte (yani cerrahi tadavi yönteminde) olumsuz tepki verenlerin sayısı

olmak üzere, test istatistiğinin örnekten hesaplanan değeri $b_h = 1$ dür.

Karar: $\alpha = 0{,}06$ önem seviyesinde H_1'e göre karar kuralı; $b_h \leq b_\alpha$ ise H_0 hipotezi ret edilir ve $b_h > b_\alpha$ ise H_0 hipotezi ret edilemez. Burada $P(b \leq b_\alpha) = \alpha$ olacak şekilde b_α kritik değeri bulunmalıdır.

b istatistiğinin örnekleme dağılımı için olasılık fonksiyonu;

$$f(x,a) = P(b = x) = \frac{\binom{n_2}{x}\binom{n_1}{a}}{\binom{n_1+n_2}{a+x}} \;, x = 0, 1, 2, \ldots, n_2 \text{ ve a=5, } a + x = 6, n_2 = 8$$

$$= \frac{\binom{8}{x}\binom{8}{a}}{\binom{16}{6}} \;; x = 0,1,2,3, \ldots, 8 \text{ ve } a + x = 6$$

dir. Buna göre; $\alpha = 0{,}06$ iken

$$P(b = 0) = \frac{\binom{8}{0}\binom{8}{6}}{\binom{16}{6}} = \frac{28}{8008} = 0{,}0035 \;; \quad P(b = 1) = \frac{\binom{8}{1}\binom{8}{5}}{\binom{16}{6}} = \frac{448}{8008} = 0{,}0559 \text{ ve}$$

$$P(b = 2) = \frac{\binom{8}{2}\binom{8}{4}}{\binom{16}{6}} = \frac{1960}{8008} = 0{,}2448$$

bulunur. Böylece;

$P(b \leq 1) = P(b = 0) + P(b = 1) = 0{,}0035 + 0{,}0559 = \mathbf{0{,}0594} < \boldsymbol{\alpha}$**=0,06…(*)**

$P(b \leq 2) = 0{,}0594 + P(b = 2) = 0{,}0594 + 0{,}2448 = 0{,}3042 > \alpha = 0{,}06$

(*) eşitsizliğine göre $b_\alpha = 1$ bulunur. Buna göre $b_h = b_\alpha = 1$ olduğundan H_0 hipotezi ret edilir. Bu sebeple tedaviye olumsuz tepki verenlerin oranının cerrahi tedavi uygulanan hastalarda daha düşük olduğu söylenebilir.

<u>Spss çözümü:</u> Değişken(X): Tedaviye tepki verme durumu…. Nitel ve ölçme düzeyi sınıflama

İlgilenilen özellik: Olumsuz tepki verme (1)

İlgilenilmeyen özellik: Olumlu tepki verme (2)

Birinci Grup: İlaçla tedavi yöntemi ve İkinci grup: Cerrahi tedavi yöntemi olup gruplar birbirinden bağımsızdır.

Π_1: İlaçla tedavi yönteminde olumsuz tepki verenlerin oranı

Π_2: Cerrahi yönteminde olumsuz tepki verenlerin oranı, olmak üzere hipotezler:

$H_0 : \Pi_1 = \Pi_2$

$H_1 : \Pi_1 > \Pi_2$ şeklinde oluşturulur.

tedavi * tepki Crosstabulation

		tepki		Total
		olumsuz	olumlu	
tedavi	ilaçla tedavi	5	3	8
	cerrahi tedavi	1	7	8
	Total	6	10	16

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4,267	1	,039		
Continuity Correction	2,400	1	,121		
Likelihood Ratio	4,557	1	,033		
Fisher's Exact Test				,119	,059
Linear-by-Linear Association	4,000	1	,046		
N of Valid Cases	16				

Test istatistiğinin örnekten hesaplanan değeri $b_h = 1$ ve H_1 tek yönlü olduğundan $p = 0{,}059$ olup, ; $\alpha = 0{,}06$ iken $p < \alpha$ dır. Bu sebeple H_0 hipotezi ret edilir. Bu sebeple tedaviye olumsuz tepki verenlerin oranının cerrahi tedavi uygulanan hastalarda daha düşük olduğu söylenebilir.