T.C.

ONDOKUZ MAYIS ÜNİVERSİTESİ

FEN EDEBİYAT FAKÜLTESİ
İSTATİSTİK BÖLÜMÜ

İST.482 PARAMETRİK OLMAYAN İSTATİSTİKSEL YÖNTEMLER

PROF. DR. YÜKSEL ÖNER

12. Hafta

özenilen üniversite

6.2.2 Çoklu Karşılaştırmalar

Kruskal – Wallis H testi sonucunda H_0 hipotezi ret edilmesi durumunda, k tane faktör düzeyinden hangilerinin farklı etkide bulunduğunu ya da k tane grubun hangilerinde ortalamaların veya medyanların farklılık gösterdiğinin belirlenmesi istenebilir. Bu amaçla kullanılan tekniklere çoklu karşılaştırma teknikleri denir. Burada bu çoklu karşılaştırma tekniklerinden Miller tarafından geliştirilmiş olan çoklu karşılaştırma tekniği verilecektir. Bu çoklu karşılaştırma tekniği iki farklı grubun karşılaştırmasını eşanlı güven aralığı kullanarak gerçekleştirmektedir. k tane grubun olduğu bir analizde mümkün olan ikili karşılaştırmaların sayısı $\binom{k}{2} = \frac{k(k-1)}{2}$ kadar olacaktır. $t < j = 1, 2, \dots, k$ olmak üzere t-nci grup ile j-nci grubu karşılaştırmak istediğimizi kabul edelim. Bu durumda test edilecek olan hipotezler:

$H_0 : \tau_t = \tau_j$

$$H_1 : \tau_t \neq \tau_j \ , t = 1, 2, \dots (k-1); j = (t+1), (t+2), \dots, k \qquad (6.19)$$

şeklindedir. Eğer; test istatistiği olarak H istatistiği kullanılmış ise

$\frac{|\bar{R}_{.t} - \bar{R}_{.j}|}{\sqrt{V(\bar{R}_{.t} - \bar{R}_{.j})}} > \sqrt{H'_\alpha}$ olduğunda $H_0 : \tau_t = \tau_j$ hipotezi ret edilir, aksi takdirde ret edilemez.

Eğer test istatistiği olarak H^* istatistiği kullanılmış ise

$\frac{|\bar{R}_{.t} - \bar{R}_{.j}|}{\sqrt{V(\bar{R}_{.t} - \bar{R}_{.j})}} > \sqrt{H^{*\prime}_\alpha}$ olduğunda $H_0 : \tau_t = \tau_j$ hipotezi ret edilir, aksi takdirde ret edilemez.

Burada $V(\bar{R}_{.t} - \bar{R}_{.j}) = \frac{n*(n+1)}{12}\left(\frac{1}{n_t} + \frac{1}{n_j}\right)$ dir.

<u>SPSS Algoritması</u>

Algoritma 1

Bu algoritma ile Spss çözümlemesinde sadece H testi uygulaması yapılır. Çoklu karşılaştırmalar yapılamaz.

Adım 1. İlgilenilen değişken (bağımlı değişken) ve faktör (grup) değişkeni özellikleri ile **Variable View** sayfasında birlikte tanımlanır. Grup değişkeninin kategorileri **Value** hücresinde kodlanır.

Adım 2. Bağımlı değişken değerleri ilgili sütuna grup sırasına göre alt alta ve grup kategorileri de kendisine ait sütuna n_1 tane "1", n_2 tane "2"… vs şeklinde girilir.

Adım 3. Analyze > Nonparametric Tests > Legacy Dialogs > K Independent Samples…yolu izlenerek açılacak olan ekranda, bağımlı değişken seçilerek **Test Variable List işlem** kutusuna ve Grup değişkeni seçilerek **Grouping Variable** işlem kutusuna aktarılır. Aktarma işleminden sonra aktif konuma gelen **Define Range** penceresi açılarak bu pencerede grup değişkeninin değer aralığı tanımlanır ve **Continue** tıklanır. Bu ekranda yer alan **Test Type**

bölümünden **Kruskal-Wallis H** işaretlenir. Yine aynı ekranda bulunan **Exact** penceresi açılarak p olasılık değerinin hesaplanmasında gerekli olan **Asymp. Sig**. seçeneği işaretlenir.

Adım 4. Continue ve **OK** düğmeleri tıklanarak test işlemi bitirilir. Sonuçlar **Output** sayfasında tablo halinde sunulur. Sonuç tablosunda test istatistiğinin hesaplanan değerini (χ_h^2) ve p olasılık değerini belirleyebiliriz. Eğer α önem seviyesi olmak üzere, eğer $p < \alpha$ ise H_0 hipotezi ret edilir, $p \geq \alpha$ ise H_0 hipotezi ret edilemez.

Algoritma 2

Bu algoritma ile Spss çözümlemisinde hem H testi uygulaması hem de çoklu karşılaştırmalar yapılabilir.

Adım 1. İlgilenilen değişken (bağımlı değişken) ve faktör (grup) değişkeni özellikleri ile **Variable View** sayfasında birlikte tanımlanır. Grup değişkeninin kategorileri **Value** hücresinde kodlanır.

Adım 2. Bağımlı değişken değerleri ilgili sütuna grup sırasına göre alt alta ve grup kategorileri de kendisine ait sütuna n_1 tane “1”, n_2 tane “2”… vs şeklinde girilir.

Adım 3. Analyze > Nonparametric Tests > Independent Samples… yolu izlenerek açılacak olan ekranda, **Fields** penceresi açılır. Bağımlı değişken seçilerek **Test Fields** işlem kutusuna ve Grup değişkeni seçilerek **Groups** işlem kutusuna aktarılır. **Settings** penceresi açılır **Customize tests** seçeneği işaretlenir ve aktif konuma gelen test seçenekleri içerisinden **Kruskal-Wallis 1-way Anova** işaretlenir. Bu işaretlemeden sonra aktif hale gelen **Multiple Comparisons:** seçenek kutusundan **All Pairwise** seçeneği seçilir

Adım 4. Run düğmesi tıklanarak test işlemi bitirilir. Sonuçlar **Output** sayfasında tablo halinde sunulur. Sonuç tablosunun üzerine çift tıklayarak detaylı sonuç tablolarına ulaşılır. Bu tablolarda hem Kruskal – Wallis H testi sonuçları hem de çoklu karşılaştırma sonuçları yer almaktadır.

Kruskal – Wallis H testi sonuç tablosunda test istatistiğinin hesaplanan değerini $(H\ veya\ \chi_h^2)$ ve p olasılık değerini belirleyebiliriz. Eğer α önem seviyesi olmak üzere, $p < \alpha$ ise H_0 hipotezi ret edilir, $p \geq \alpha$ ise H_0 hipotezi ret edilemez.

Çoklu karşılaştırma sonuç tablosuna ulaşabilmek için ekran sayfasının son satırında yer alan **View: Independent Samples Test View** hücresi tıklanır ve açılan seçeneklerden **Pairwise Comparisons** seçeneği seçilir. Ekrana gelen grafik ve tablo değerlendirilir. Tablo renkli işaretli satır/satırlarda yer alan gruplar arasında anlamlı farklılık vardır (yani $p \leq \alpha$). Diğer satırlarda yer alan gruplar arasında anlamlı farklılık yoktur (yani $p > \alpha$).

ÖRNEK 6.2 $k = 3$ ve $n_1 = 2, n_2 = 1, n_3 = 1$ iken H istatistiğinin;

a) örnekleme dağılımını oluşturunuz?

b) beklenen değe ve varyansını hesaplayınız?

Çözüm **a)** $H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_{.j}^2}{n_j} - 3(n+1)$

H istatistiğinin hesaplanabileceği farklı durumların sayısı $\binom{n}{n_1,\ n_2,\ \dots\ ,\ n_k}$ multinomiali (çok terimlisi) olup;

$$\binom{n}{n_1,\ n_2,\ \dots\ ,\ n_k} = \binom{n}{n_1}\binom{n-n_1}{n_2}\dots\binom{n-n_1-n_2-\dots-n_{k-1}}{n_k} \qquad (6.20)$$

dir. Eşitlik (6.20) dikkate alındığında $k = 3$ ve $n_1 = 2$, $n_2 = 1$, $n_3 = 1$ iken H istatistiğinin hesaplanabileceği farklı durumların sayısı; $n = \sum_{j\_1}^{3} n_j = 4$ olmak üzere

$\binom{n}{n_1,\ n_2,\ n_3} = \binom{4}{2,\ 1,\ 1} = \binom{4}{2}\binom{2}{1}\binom{1}{1} = 12$ dir.

Durum	R_{i1}	R_{i2}	R_{i3}	$R_{.1}$	$R_{.2}$	$R_{.3}$	$H = \frac{12}{4*5}\sum_{j=1}^{3}\frac{R_{.j}^2}{n_j} - 3(n+1)$
1	1, 2	3	4	3	3	4	2,7
2	1, 2	4	3	3	4	3	2,7
3	1, 3	2	4	4	2	4	1,8
4	1, 3	4	2	4	4	2	1,8
5	1, 4	2	3	5	2	3	0,3
6	1, 4	3	2	5	3	2	0,3
7	2, 3	1	4	5	1	4	2,7
8	2, 3	4	1	5	4	1	2,7
9	2, 4	1	3	6	1	3	1,8
10	2, 4	3	1	6	3	1	1,8
11	3, 4	1	2	7	1	2	2,7
12	3, 4	2	1	7	2	1	2,7

H istatistiğinin örnekleme dağılımı

$H = h$	0,3	1,8	2,7
$\Pr(H = h)$	2/12	4/12	6/12

b) H istatistiğinin beklenen değeri; $E(H) = \sum_h h\Pr(H = h) = \frac{1}{12}(0{,}6 + 7{,}2 + 16{,}2) = 2$

H istatistiğinin varyansı; $V(H) = E(H^2) - [E(H)]^2 = \sum_h h^2 \Pr(H = h) - (2)^2$

$= \frac{1}{12}(0{,}18 + 12{,}96 + 43{,}74) - 4 = 0{,}74$ bulunur.

ÖRNEK 6.3 $k = 3$ ve $n_1 = 2$, $n_2 = 2$, $n_3 = 1$ iken H istatistiğinin;

a) örnekleme dağılımını oluşturunuz?

b) beklenen değe ve varyansını hesaplayınız?

Çözüm **a)** $H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_{.j}^2}{n_j} - 3(n+1)$

H istatistiğinin hesaplanabileceği farklı durumların sayısı Eşitlik (6.20) dikkate alındığında $k = 3$ ve $n_1 = 2$, $n_2 = 2$, $n_3 = 1$ iken $n = \sum_{j\_1}^{3} n_j = 5$ olmak üzere

$\binom{n}{n_1,\ n_2,\ n_3} = \binom{5}{2,\ 2,\ 1} = \binom{5}{2}\binom{3}{2}\binom{1}{1} = 30$ dur. $3(n+1) = 18$ olduğu dikkate alınır.

Durum	R_{i1}	R_{i2}	R_{i3}	$R_{.1}$	$R_{.2}$	$R_{.3}$	$H = \frac{12}{5*6}\sum_{j=1}^{3}\frac{R_{.j}^2}{n_j} - 18$
1	1, 2	3, 4	5	3	7	5	3,6
2	1, 2	3, 5	4	3	8	4	3,0
3	1, 2	4, 5	3	3	9	3	3,6
4	1, 3	2, 4	5	4	6	5	2,4
5	1, 3	2, 5	4	4	7	4	1,4
6	1, 3	4, 5	2	4	9	2	3,0
7	1, 4	2, 3	5	5	5	5	2,0
8	1, 4	2, 5	3	5	7	3	0,4
9	1, 4	3, 5	2	5	8	2	1,4
10	1, 5	2, 3	4	6	5	4	0,6
11	1, 5	2, 4	3	6	6	3	0,0
12	1, 5	3, 4	2	6	7	2	0,6
13	2, 3	1, 4	5	5	5	5	2,0
14	2, 3	1, 5	4	5	6	4	0,6
15	2, 3	4, 5	1	5	9	1	3,6
16	2, 4	1, 3	5	6	4	5	2,4
17	2, 4	1, 5	3	6	6	3	0,0
18	2, 4	3, 5	1	6	8	1	2,4
19	2, 5	1, 3	4	7	4	4	1,4
20	2, 5	1, 4	3	7	4	3	0,4
21	2, 5	3, 4	1	7	7	1	2,0
22	3, 4	1, 2	5	7	3	5	3,6
23	3, 4	1, 5	2	7	6	2	0,6
24	3, 4	2, 5	1	7	7	1	2,0
25	3, 5	1, 2	4	8	3	4	3,0
26	3, 5	1, 4	2	8	5	2	1,4
27	3, 5	2, 4	1	8	6	1	2,4
28	4, 5	1, 2	3	9	3	3	3,6
29	4, 5	1, 3	2	9	4	2	3,0
30	4, 5	2, 3	1	9	5	1	3,6

H istatistiğinin örnekleme dağılımı

$H = h$	0,0	0,4	0,6	1,4	2,0	2,4	3,0	3,6
$\Pr(H = h)$	2/30	2/30	4/30	4/30	4/30	4/30	4/30	6/30

b) H istatistiğinin beklenen değeri; $E(H) = \sum_h h\Pr(H = h)$

$$= \frac{1}{30}(0 + 0{,}8 + 2{,}4 + 5{,}6 + 8{,}0 + 9{,}6 + 12{,}0 + 21{,}6) = 2$$

H istatistiğinin varyansı; $V(H) = E(H^2) - [E(H)]^2 = \sum_h h^2 \Pr(H = h) - (2)^2$

$= \frac{1}{30}(0 + 0{,}32 + 1{,}44 + 7{,}84 + 16 + 23{,}04 + 36 + 77{,}76) - 4 = 1{,}413$ bulunur.

ÖRNEK 6.4 Rastgele seçilen 12 hasta yine rastgele olarak 3 gruba ayrılmıştır. Her bir gruptaki hastalar üzerinde farklı bir ilaç tedavisi uygulanmış ve hastaların iyileşme süresi ölçülmüştür. Bu verilere göre;

a) ilaç türünün iyileşme süresi üzerinde etkisinin farklı olup olmadığına %5 önem seviyesinde karar veriniz?

b) ilaç türünün etkisi farklı ise hangi ilaç türleri arasında farklılık olduğunu belirleyiniz?

İLAÇ TÜRÜ			
A	**B**	**C**	
12 (4)	24 (9)	23 (8)	
10 (3)	18 (7)	30 (10)	
7 (2)	16 (6)	31 (11)	
6 (1)	15 (5)	32 (12)	
$n_1 = 4$	$n_2 = 4$	$n_3 = 4$	n=12
$R_{.1}$=10	$R_{.2}$=27	$R_{.3}$=41	$R_{..}$=78

Çözüm: Bağımlı değişken (X): İyileşme süresi… Nicel, sürekli ve ölçme düzeyi oranlama

Faktör: İlaç Türü… Nitel ve ölçme düzeyi sınıflama

Faktör düzeyleri $\left\{\begin{array}{l} A \text{ İlaç türü } (1) \\ B \text{ İlaç türü } (2) \\ C \text{ İlaç Türü } (3) \end{array}\right\}$ Bağımsız gruplar $(k = 3)$

a) İlaç türünün iyileşme süresi üzerinde etkisinin farklılığı ile ilgili **hipotezler:**

$H_0 : \tau_1 = \tau_2 = \tau_3$

$H_1 : \tau_j$'lerden en az biri farklıdır.

Test istatistiği: Veriler içerisinde aynı değerli gözlemler bulunmadığından H istatistiğidir. H istatistiğinin alabileceği değeri hesaplamak için üç grup birleştirilerek oluşturulan birleştirilmiş örnekte örnek birimlerine verilen sıra sayıları ve gruplara göre bu sıra sayılarının toplamları tablo üzerinde gösterilmiştir.

$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_{.j}^2}{n_j} - 3(n+1) \Rightarrow$ Test istatistiğinin örnekten hesaplanan değeri $H_h = \frac{12}{12*13}\left[\frac{10^2}{4} + \frac{27^2}{4} + \frac{41^2}{4}\right] - 3*13$ =9,269 olarak bulunur.

Karar: Karar kuralı; α önem seviyesinde kritik değer H'_α olmak üzere $H_h \geq H'_\alpha$ ise H_0 hipotezi ret edilir, aksi takdirde ret edilemez. Veya $\Pr(H \geq H_h) = p$ olmak üzere $p \leq \alpha$ ise H_0 hipotezi ret edilir, $p > \alpha$ ise ret edilmez.

Grup sayısı $k = 3$ ve $\forall n_j < 5$ olduğundan kritik değer bulunmasında H istatistiğinin kesin örnekleme dağılımından yararlanılır ve (T15) tablosu kullanılır. $\alpha = 0{,}05$; $k = 3$ ve $n_1 = n_2 = n_3 = 4$ olduğundan kritik değer $H'_\alpha = 5{,}692$ dir. 9,269 > 5,692 yani $H_h > H'_\alpha$ olduğundan H_0 hipotezi ret edilir. ($p = Pr(H \geq H_h) = Pr(H \geq 9{,}269)$ =? T15 tablosundan $k = 3$ ve $n_1 = n_2 = n_3 = 4$ olduğunda $Pr(H \geq 7{,}654) = 0{,}008$ olup, buna göre $p = Pr(H \geq 9{,}269) < 0{,}008$ ve böylece $p < \alpha = 0{,}05$ olacağından H_0 hipotezi ret edilir.)

Sonuç olarak ilaç türlerinden en az birisi hastaların iyileşme süresi üzerinde farklı etki göstermektedir.

b) Hangi ilaç türlerinin farklı etki yaptığını belirlemek için çoklu karşılaştırma tekniği uygulanır. Mümkün olan ikili karşılaştırmaların sayısı $\binom{k}{2} = \binom{3}{2} = 3$ tanedir.

<u>A(1) ilacı ile B(2) ilacı için</u>

Hipotezler

$H_0: \tau_1 = \tau_2 \quad \overline{R}_{.1} = \frac{R_{.1}}{n_1} = \frac{10}{4} = 2{,}5$; $\overline{R}_{.2} = \frac{R_{.2}}{n_2} = \frac{27}{4} = 6{,}75$;

$H_1: \tau_1 \neq \tau_2 \quad V(\overline{R}_{.1} - \overline{R}_{.2}) = \frac{n*(n+1)}{12}\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = \frac{12*13}{12}\left(\frac{1}{4} + \frac{1}{4}\right) = 6{,}5$; $H'_\alpha = 5{,}692$

$\frac{|\overline{R}_{.1} - \overline{R}_{.2}|}{\sqrt{V(\overline{R}_{.1} - \overline{R}_{.2})}} > \sqrt{H'_\alpha}$ ise H_0 hipotezi ret edilir, aksi takdirde kabul edilir.

$\frac{|\overline{R}_{.1} - \overline{R}_{.2}|}{\sqrt{V(\overline{R}_{.1} - \overline{R}_{.2})}} = \frac{|2{,}5 - 6{,}75|}{\sqrt{6{,}5}} = 1{,}667$ ve $\sqrt{H'_\alpha} = \sqrt{5{,}692} = 2{,}386$ olup, 1,667 < 2,386 olduğundan H_0 hipotezi ret edilemez. Buna göre A ve B ilaç türlerinin iyileşme süresine etkileri aynıdır.

<u>A(1) ilacı ile C(3) ilacı için</u>

Hipotezler

$H_0: \tau_1 = \tau_3 \quad \overline{R}_{.1} = \frac{R_{.1}}{n_1} = \frac{10}{4} = 2{,}5$; $\overline{R}_{.3} = \frac{R_{.3}}{n_3} = \frac{41}{4} = 10{,}25$;

$H_1: \tau_1 \neq \tau_3 \quad V(\overline{R}_{.1} - \overline{R}_{.3}) = \frac{n*(n+1)}{12}\left(\frac{1}{n_1} + \frac{1}{n_3}\right) = \frac{12*13}{12}\left(\frac{1}{4} + \frac{1}{4}\right) = 6{,}5$; $H'_\alpha = 5{,}692$

$\frac{|\overline{R}_{.1}-\overline{R}_{.3}|}{\sqrt{V(\overline{R}_{.1}-\overline{R}_{.3})}} > \sqrt{H'_\alpha}$ ise H_0 hipotezi ret edilir, aksi takdirde kabul edilir.

$\frac{|\overline{R}_{.1}-\overline{R}_{.3}|}{\sqrt{V(\overline{R}_{.1}-\overline{R}_{.3})}} = \frac{|2,5-10,25|}{\sqrt{6,5}} = 3,04$ ve $\sqrt{H'_\alpha} = \sqrt{5,692} = 2,386$ olup, 3,04 > 2,386 olduğundan H_0 hipotezi ret edilir. Buna göre A ve C ilaç türlerinin iyileşme süresine etkileri farklıdır. Üstelik A ilacı kullananlar daha kısa sürede iyileşme göstermektedirler.

<u>B(2) ilacı ile C(3) ilacı için</u>

Hipotezler

$H_0 : \tau_2 = \tau_3 \quad \overline{R}_{.2} = \frac{R_{.2}}{n_2} = \frac{27}{4} = 6,75;\ \overline{R}_{.3} = \frac{R_{.3}}{n_3} = \frac{41}{4} = 10,25$

$H_1 : \tau_2 \neq \tau_3 \quad V(\overline{R}_{.2} - \overline{R}_{.3}) = \frac{n*(n+1)}{12}\left(\frac{1}{n_2}+\frac{1}{n_3}\right) = \frac{12*13}{12}\left(\frac{1}{4}+\frac{1}{4}\right) = 6,5\ ;\ H'_\alpha = 5,692$

$\frac{|\overline{R}_{.2}-\overline{R}_{.3}|}{\sqrt{V(\overline{R}_{.2}-\overline{R}_{.3})}} > \sqrt{H'_\alpha}$ ise H_0 hipotezi ret edilir, aksi takdirde kabul edilir.

$\frac{|\overline{R}_{.2}-\overline{R}_{.3}|}{\sqrt{V(\overline{R}_{.2}-\overline{R}_{.3})}} = \frac{|6,75-10,25|}{\sqrt{6,5}} = 1,374$ ve $\sqrt{H'_\alpha} = \sqrt{5,692} = 2,386$ olup, 1,374 < 2,386 olduğundan H_0 hipotezi ret edilemez. Buna göre B ve C ilaç türlerinin iyileşme süresine etkileri aynıdır.

<u>Spss Çözümü</u>

Algoritma 1'e göre:

Ranks

	ilaç	N	Mean Rank
süre	A ilacı	4	2,50
	B ilacı	4	6,75
	C ilacı	4	10,25
	Total	12	

Test Statistics[a,b]

	süre
Chi-Square	9,269
Df	2
Asymp. Sig.	,010

$p = 0,01 < \alpha = 0,05$ olduğundan H_0 hipotezi ret edilir. Sonuç olarak ilaç türlerinden en az birisi hastaların iyileşme süresi üzerinde farklı etki göstermektedir.

Algoritma 2’ye göre:

Total n	12	$2p = 0{,}01$ olup $p = \frac{0{,}01}{2} = 0{,}005$ ve böylece $p = 0{,}005 < \alpha = 0{,}05$ olduğundan H_0 hipotezi ret edilir. Sonuç olarak ilaç türlerinden en az birisi hastaların iyileşme süresi üzerinde farklı etki göstermektedir.
Test Statistic	9,269	
Degrees of freedom	2	
Asymptotic Sig. (2-sided test)	0,010	

Saple 1-Sample 2	Test statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig. (p)	
A – B	-4,250	2,55	-1,667	0,096	0,287	$p > \alpha$; $H_0: \tau_1 = \tau_2$ Kabul
A – C	-7,750	2,55	-3,040	0,002	0,007	$p < \alpha$; $H_0: \tau_1 = \tau_3$ Ret
B – C	-3,500	2,55	-1,373	0,170	0,509	$p > \alpha$; $H_0: \tau_2 = \tau_3$ Kabul