Bilişimde İstatistiksel Analiz

Ders 12

Ali Mertcan KOSE Msc.
amertcankose@ticaret.edu.tr

İstanbul Ticaret Üniversitesi

İSTANBUL TİCARET ÜNİVERSİTESİ

Bağımlı İki Grupların Karşılaştırılması

- *Parametrik test:* Bağımlı İki örneklem t Testi
- *Parametrik olmayan test:* Wilcoxon Testi

Bağımlı İki Örneklem t Testi

Parametrik test varsayımları sağlandığında, ölçümle belirtilen sürekli bir değişken yönünden aynı bireylerin değişik iki zaman ya da durumdaki ölçümleri arasında fark olup olmadığını test etmek için kullanılır.

Bağımlı İki Örneklem t Testi

Dikkat edilmesi gereken noktalar:

- Veri ölçümle belirtilmiştir.
- Aynı bireyler üzerinde aynı konuda iki kez ölçüm yapılmaktadır.

Varsayımları:

İki grup arasındaki değerlere ilişkin fark değerleri dağılımının normal dağılım göstermesi gerekir.

Note

Varsayım sağlanmıyor ise; Bu testin Parametrik olmayan karşılığı olan Wilcoxon iki örneklem testi kullanılmalıdır.

Bağımlı İki Örneklem t Testi

İki bağımlı örneklem arasındaki farkın test edildiği durumları Üç grupta toplanabilir

1. Ölçümle belirtilen bir değişken yönünden aynı bireylerin değişik iki zaman ya da durumdaki ölçümlerinin farklı olup olmadığının test edilmesinde kullanılır.

Örnek

Kandaki şeker miktarını düşürmek için hazırlanan bir diyet programının etkinliğini ölçmek için şeker hastalarının diyetten önce kandaki şeker miktarları ile diyetten sonraki kandaki şeker miktarlarının farklı olup olmadığını test etmek için kullnılır.

2. Değişik iki ölçüm aracının aynı bireylerde aynı ölçümü yapıp yapmadığını ya da aynı sonucu verip vermediğini test etmek için kullanılır.

Örnek:

İki ayrı firmanın ürettiği tansiyon ölçme araçlarının aynı kişilerin tansiyonunu aynı değerde ölçüp ölçmediğinin test edilmesinde kullanılır.

Bağımlı İki Örneklem t Testi

3. Değişik iki ölçümcünün aynı ölçüm aracıyla aynı bireylerin ölçümünü aynı değerde yapıp yapmadıklarının (ölçümcü farklılıklarının) test edilmesinde kullanılır.

Örnek:

Biri uzman, diğeri acemi olan iki ölçücünün bireylerin vücut yağ yüzdelerini deri kıvrımı kalınlığı yöntemiyle ölçümleri için kullanılır.

Hipotezlerin Kurulması

$H_0 : \mu_1 - \mu_2(\bar{D}) = 0 \rightarrow$ iki bağımlı değişken arasında fark yoktur.

$H_1 : \mu_1 - \mu_2(\bar{D}) \neq 0 \rightarrow$ iki bağımlı değişken arasında fark vardır.

$H_1 : \mu_1 - \mu_2(\bar{D}) > 0 \rightarrow$ iki bağımlı değişken arasında uygulama öncesi sonrasından büyüktür.

$H_1 : \mu_1 - \mu_2(\bar{D}) < 0 \rightarrow$ iki bağımlı değişken arasında uygulama öncesi sonrasından küçüktür.

Bağımlı İki Örneklem t Testi

Test İstatistiğinin hesaplanması

- Gözlemlerin önceki değerlerinden sonraki değerleri çıkartılarak fark dizisi oluşturulur.
- Farkların Ortalması bulunur: $\bar{D}$
- Farkların standart sapması bulunur: S_D
- Farkların standart hatası bulunur: $S_D = \frac{S_D}{\sqrt{n}}$
- Test istatistiği (t_h) hesaplanır. $t = \frac{\bar{D}}{S_{\bar{D}}}$

Bağımlı İki Örneklem t Testi

α anlamlılık düzeyinin belirlenmesi

İstatistiksel Karar

- Bulunan t_{hesap} istatistiği, seçilen α anlamlılık düzeyi ve n-1 serbestlik derecesindeki t_{tablo} istatistiği ile karşılaştırılır.

$|t_{hesap}| > t_{tablo}$ ise “iki bağımlı değişken arasında fark yoktur.” şeklinde kurukan H_0 hipotezi reddedilir ve p<α yazılır.

Bağımlı İki Örneklem t Testi

Güven Aralığı

$\bar{D} \pm t_{(\alpha/2,n-1)} \frac{S_D}{\sqrt{n}}$

$\mathrm{P}(\bar{D} - t_{t(\alpha/2,n-1)} \frac{S_D}{\sqrt{n}} \leq \mu_D \leq \bar{D} + t_{t(\alpha/2,n-1)} \frac{S_D}{\sqrt{n}})$

Örnek 1

Sigara içenler arasından rasgele olarak seçilen 10 kişinin günde ne kadar sigara içtikleri saptanmıştır.(X_{1i} :işlem öncesi) Sonra aynı kişilere bir sağlık uzmanı sağlık üzerine etkilerini konu edinen bir seminer vermiştir. Seminerden sonra aynı kişilerin günde ne kadar sigara içtikleri saptanmıştır.(X_{2i} :işlem sonrası)

X_{1i} : 30,25,25,20,20,18,17,17,15,13

X_{2i} : 28,25,25,18,17,18,16,16,15,12

$\alpha = 0.05$ anlamlılık düzeyinde seminerin etkili olduğu söylenebilir mi? %95 güven düzeyinde güven aralığını oluşturunuz.

Örnek 1

x1	x2	Di
30	28	2
25	25	0
25	25	0
20	18	2
20	17	3
18	18	0
17	16	1
17	16	1
15	15	0
13	12	1

Örnek 1

$H_0 : \mu_D = 0$

$H_1 : \mu_D > 0$

$t_h = \frac{\bar{D} - \mu_D}{S_D/\sqrt{n}}$

$\sum_{n=1}^{n} D_i = 10, \ \sum_{n=1}^{n} D_i^2 = 20, \ \bar{D} = \frac{\sum_{n=1}^{10} D_i}{10} = \frac{10}{10} = 1$

$S_D = \sqrt{\frac{\sum_{n=1}^{10} D_i^2 \frac{(\sum_{n=1}^{10} D_i^2)}{n}}{n-1}} = \sqrt{\frac{20 - \frac{10^2}{10}}{10-1}} = \frac{\sqrt{10}}{3}$

$t_H = \frac{\bar{D} - \mu_D}{S_D/\sqrt{n}} = \frac{1}{\frac{\sqrt{10}}{3}/\sqrt{10}} = 3$

$t_H > T_{T(0.05,9)} = 3 > 1.833$ olduğundan H_0 reddedilir.. Yani bağımlı iki örneklem arasında fark vardır. İçilen sigara miktarının seminerden sonra azaldığı %95 güvenle söylenebilir.

Örnek 1

Güven Aralığı

$P(\bar{D} - t_{t(\alpha/2,n-1)}\frac{S_D}{\sqrt{n}} \leq \mu_D \leq \bar{D} + t_{t(\alpha/2,n-1)}\frac{S_D}{\sqrt{n}})$

$\bar{D} \pm t_{(\alpha/2,n-1)}\frac{S_D}{\sqrt{n}} \Rightarrow 1 \pm 2.262\times\ \frac{1}{3}$ [0.146,1.754]

Bu aralığın μ_D'yi içeren aralıklardan biri olması olasılığı %95 tir.

Wilcoxon Bağımlı İki Örneklem Testi

Test istatistiğinin hesaplanması incelenen denek sayısının 25'ten az olup olmama durumuna göre ayrı işlemlerle yapılır.

- Denek sayısı 25'ten az olduğunda test işlemleri
  - Her kişinin değerleri önce ve sonra kolonlarına yazılır.
  - İki ölçüm arasındaki farklar (önce-sonra) alınır ve fark kolonuna yazılır. Fark değerlerine işaret dikkate alınmadan küçükten büyüğe doğru sıra numarası verilir ve sıra no sütunu elde edilir.

Wilcoxon Bağımlı İki Örneklem Testi

- Fark dizininde sıfır değerini alan fark ya da farklar var ise aşağıdaki kurallar uygulanır.

1. Fark konumunda bir tane sıfır var ise: Bu değer değelendirmeden çıkartılır ve denek sayısı bir azaltılır.
2. Fark konumundaki sıfır sayısı çift ise: Önce sıfırlar sıralanır. Sıfıra karşılık gelen sıra numaralarının ortalaması sıfırların sıra numarası olur sıfırların sıra numarasının yarısına +, yarısına - işareti konur.

Wilcoxon Bağımlı İki Örneklem Testi

3. Fark konmunundaki sıfır sayısı tek ise: Sıfırın herhangi bir tanesi değerlendirmeden çıkartılır. Denek sayısı bir azaltılır. Sıra numarası verme ve işaretleme işlemi 2. maddesindeki gibi yapılır. + Fark konumunda sıfırlar ve aynı değeri alan gözlemler var ise "yeni sıra no kolonu" oluşturulur. + Farkların işaretleri sıra numaralarının önüne yazılır ve "işaretli yeni sıra no" sütünu oluşturulur. + Test istatistiği'nin elde edilmesi: Farklara ilişkin işaretli sıra numaralarından sayısı az olan işaretin sıra numaraları toplanır T istatisiği elde edilir.

Wilcoxon Bağımlı İki Örneklem Testi

İstatistiksel karar Hesapla bulunan t değeri t_{tablo} değerinden küçükse H_0 hipotezi reddedilir.

- Denek Sayısı 25 ya da 25'den fazla olduğunda test işlemleri

z istatistiğinden yararlanılır.

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

Burada,

T: A maddesinde bulunan T hesap istatistiği

n: Gözlem sayısı

Wilcoxon Bağımlı İki Örneklem Testi

İstastiksel Karar

Z değerine ilişkin olasılık z tablosundan bulunur ve 0.5'den çıkartılır.

H_1 hipotezi tek yönlü ise tablo olasılık değeri ile önceden belirlenen α anlamlılık düzeyinde doğrudan karşılaştırılır.

H_1 hipotezi çift yönlü ise tablo olasılık değeri 2 ile çarpıldıktan sonra önceden belirlenen α anlamlılık düzeyi ile karşılaştırılır.

Tablo olasılık değeri önceden saptanan α anlamlılık düzeyinden küçük olursa H_0 hipotezi reddedilir.

Örnek 2

12 bireyin diyet öncesi ağırlıklarının diyet sonrasında değişip değişmediği incelenmek isteniyor

Hipotezler:

H_0: iki bağımlı Örneklem arasında fark yoktur.

H_1: iki bağımlı Örneklem arasında fark vardır.

Örnek 2

Önce	Sonra	Fark	Sıralı fark	Sıra no	Yeni sıra no	İşaretli yeni sıra no
62	63	-1	0	1	1,5	-1,5
55	50	5	0	2	1,5	1,5
68	65	3	1	3	4,5	4,5
56	54	2	1	4	4,5	4,5
78	70	8	-1	5	4,5	-4,5
51	51	0	-1	6	4,5	-4,5
56	54	2	2	7	7,5	7,5
61	60	1	2	8	7,5	7,5
66	62	4	3	9	9	9
51	50	1	4	10	10	10
54	55	-1	5	11	11	11
61	61	0	8	12	12	12

Figure 1: Wilcoxon Test İstatistiği için Hazırlık İşlemleri Tablosu.

Örnek 2

İşaretli yeni sıra no sütunundan + ve - işaretlerinden az olanların sıra numaraları toplamıdr.

Buna göre:

t_h= 1.5+4.5+4.5 = 10.5

t_h =10.5 $<$ $t_{t(\alpha)}$=14 p<0.05

Buna göre Diyetten sonra bireylerin ağırlıklarındaki değişim istatistiksel olaran anlamlıdır.

Örnek 2

	α			
Tek kuyruk	.05	.025	.01	.005
Çift kuyruk	.10	.05	.02	.01
n				
5	1	-	-	-
6	2	1	-	-
7	4	2	0	-
8	6	4	2	0
9	8	6	3	2
10	11	8	5	3
11	14	11	7	5
12	17	14	10	7
13	21	17	13	10
14	26	21	16	13
15	30	25	20	16
16	36	30	24	19
17	41	35	28	23
18	47	40	33	28
19	54	46	38	32
20	60	52	43	37
21	68	59	49	43
22	75	66	56	49
23	83	73	62	55
24	92	81	69	61
25	101	90	77	68
26	110	98	85	76
27	120	107	93	84
28	130	117	102	92
29	141	127	111	100
30	152	137	120	109

Figure 2: Wilcoxon İki Örnek Testi Tablosu.

Örnek 2

Aynı örneğin, birey sayısı 25'in üzerindeymiş gibi düşünülüp z değeri yardımıyla çözümmğ:

$$Z= \frac{11-\frac{12(12+1)}{4}}{\sqrt{\frac{12(12+1)((2*12)+1)}{24}}}=2.20$$

p= 2*(0.5-0.4861) = 0.0278

$p=0.0278 < 0.05$ olmasından dolayı iki bağımlı örneklem arasında fark vardır.

ÖDEV 1

Kitap okumanın göz içi basıncı arttırdığı düşünülmektedir. Göz içi basınç değerlerinin normal dağılıma sahip olduğu bilinmektedir. Kitap okuduktan sonraki göz içi basıncı anlamlı düzeyde artış göstermiş midir? ($\alpha = 0.05$ anlamlılık düzeyinde hipotezi test edin ve güven aralığını belirleyiniz)

ÖDEV 1

Önce göz içi basıncı	Sonra Göz içi basıncı
10	13
15	17
18	22
10	15
16	18

ÖDEV 2

12 hatanın ilaç alımından önce ve ilaç alımından sonraki siatolik kan basıncı değerleri aşağıdaki gibidir. İlaç sistolik kan basıncını düşürmüş müdür? ($\alpha = 0.05$ anlamlılık düzeyinde test ediniz)

Önce: 14,15,17,13,14,15,16,14,14,15,14,16

sonra:13,14,15,13,15,12,14,11,10,14,16,10