

Kontenjans Tabloları

Nitel verileri iki nitel değişkene göre sınıflandırdığımızda istatistikte **çift girişimli verileri** (cross tabulated data) bize gösteren tabloları elde ederiz. Bu tablolara **çapraz tablolar** (cross tables) veya **kontenjans tablosu** (contingency tables) adı verilir. Örneğin izleyen Tabloda bir üniversiteye başvuran kişilerin, cinsiyet ve başvurunun kabulü nitel değişkenlerine göre böyle bir çapraz tablo şeklindeki dağılımı görülmektedir.

	Kabul	Red	Toplam
Erkek	233 (%42)	324 (%58)	557
Kız	88 (%31)	194 (%69)	282
Toplam	321	518	839

Bu noktaya kadar sözünü ettiğimiz frekanslar, mutlak frekanslardır. Frekanslar negatif olamazlar. Mutlak frekanslar pozitif olmak kaydıyla herhangi bir sayı olabilirler ve sınıfların frekanslarının toplamı bize toplam frekansı verir. Bir sınıfın **relatif frekansı** (relative frequency) ise, o sınıfın frekansının toplam frekansa bölünmesiyle bulunur. Bir olasılık tanımı olan **relatif frekans** yaklaşımına olasılıkla ilgili bölümde de değineceğiz. Relatif frekansların değerlere nasıl dağıldığını gösteren dağılımlar **relatif frekans dağılımı** (relative frequency distribution) adını alırlar.

x_i	n_i	Relatif Frekans
3	1	1/20
4	3	3/20
5	1	1/20
7	1	1/20
1	4	4/20
1	2	2/20
8	1	1/20
2	1	1/20
3	1	1/20
2	1	1/20
5	1	1/20
2	2	2/20
7	1	1/20
3	1	1/20
7	1	1/20
4	1	1/20
5	1	1/20
5	2	2/20
6	1	1/20
	20	20/20=1

Simpson paradoksu

Gruplama

Bir seride birbirine yakın olan deęerleri gruplar halinde toplama iřlemine **gruplama** (grouping) adı verilir. Elimizdeki veriler sınıflama ile daha derli toplu ve özet bir hale getirilmiřken, gruplama iřlemi verileri daha da özet bir hale koyar.

Aynı verileri kullanarak bu kez de 3 sınıflı bir gruplanmış seri oluşturacağımızı varsayalım. Basit serideki en büyük deęer ile en küçük deęer arasındaki farkı bularak oluşmasını istediğimiz sınıf sayısına böldüğümüzde: $(10-1)/3=3$

3 deęeri **sınıf aralığı** (class interval) olarak bulunur. Basit serinin en küçük deęeri olan 1 deęerine sınıf aralığı olarak belirlediğimiz 3 deęerini ekleyerek, gruplamadaki birinci sınıfımızın üst sınırını elde ederiz ve aynı iřlemi sürdürürüz.

Sınıfla r	f
1-4	13
5-8	18
9-12	4
	35

Sınıflama (classification), basit serideki deęerlerin birer kez ve bunların karşılıklarına ise yığında tekrar sayıları olan frekansların yazılmasıyla yapılır. Bu iřlemin sonucunda elde edilen seriye **sınıflanmış seri** adı verilir. Az önce elde ettiğimiz basit seri sınıflandırılırsa, yandaki sınıflanmış seri elde edilir. Bulunan sınıflanmış serinin frekanslarını incelediğimizde en çok tekrarlanan deęerin 6 frekansıyla 5 deęeri olduğunu, ikinci en büyük frekansın 5 olduğunu görmekteyiz. Basit seride 35 terim söz konusu olduğu için bu seri sınıflandığında toplam frekans yine 35 olmuřtur.

x	f
1	1
2	3
3	4
4	5
5	6
6	5
7	4
8	3
9	2
10	2

Çok sayıda veriyi özetlemek, grafiklerle görselleştirmek ve elde ettiği bilgiyi iletmekle ilgilenen istatistik için bir veri kümesinin hangi değere eğilim gösterdiğini belirlemek önemlidir. Frekans dağılımları ile uğraşırken, birçok veri grubundaki gözlenen değerlerin belirli değerlere yönelme eğiliminde olduklarını görürüz. Veriler arasında bazı değerlerin diğerlerine göre daha çok tekrarlanmaları, daha çok gözlenmeleri onların bu eğilimlerini açık bir şekilde ortaya koyar. İşte bu tür değerlere biz ortalama (mean) adını veriyoruz. Bütün ortalamalar, hesaplandıkları veri kümesinde bulunan en küçük gözlem değerinden daha küçük ve en büyük gözlem değerinden daha büyük olamazlar.

KURAL: Bütün ortalamalar bir veri grubundaki en küçük değerden daha küçük veya en büyük değerden daha büyük olamazlar.

İSTATİSTİKTE İKİ TÜR ORTALAMA SÖZ KONUSUDUR:

- 1- Verilerin tüm değerlerinden yararlanılarak hesaplanan ortalamalar
- 2- Verilerin bazı değerlerinden yararlanılarak hesaplanan ortalamalar

İlk gruba giren aritmetik ortalamalar, tartılı ortalamalar, kareli ortalamalar, harmonik ortalamalar ve geometrik ortalamalarda veri kümesindeki tüm verileri hesaplamalara alınız. Buna karşılık, mod, medyan ve kartiller gibi ortalamaların hesaplanmasında bir veri kümesinin sadece bazı değerleri hesaplamalarda kullanılır, değerlerin çoğu hesaplamalara alınmaz.

Mod tüm ölçeklerin verileri ile medyan ise en az sıralayıcı veya daha yüksek düzeyli ölçeklerin verileri ile hesaplanabilir. Aritmetik ortalama için ise, en az eşit aralıklı veya daha yüksek düzeydeki ölçek verileri gereklidir.

3.1. Aritmetik ortalama

İstatistikteki aritmetik ortalama (arithmetic mean) kavramı fizikteki ağırlık merkezi kavramına benzer. Nasıl fizikte bir kütlenin ağırlık merkezi hesaplanabiliyorsa, istatistikte de bir veri grubunda yer alan gözlemlerin daha çok hangi değere eğilimli oldukları aritmetik ortalama ile belirlenebilir. Aritmetik ortalama bulmak için veri kümesinde bulunan tüm ölçümlerin toplanarak veri sayısına bölünmesi gerekir.

Aritmetik ortalama, veri kümesinde bulunan tüm ölçümlerin toplanarak veri sayısına bölünmesi ile bulunur:

$$\bar{x} = \frac{\sum x_i}{n}$$

ARİTMETİK ORTALAMANIN ÖZELLİKLERİ:

- En çok bilinen ve en yaygın olarak kullanılan bir ortalamadır.
- Aritmetik ortalamanın değeri her gözlemin değerinden etkilenir.
- Bu nedenle, aykırı ve uç değerlere karşı çok duyarlıdır.
- Açık uçlu dağılımlarda hesaplanamaz.
- Anakütleye ilişkin tahmin gibi çıkarımların yapılmasında en güvenilir ortalamadır.
- Aritmetik ortalamadan sapmaların toplamı sıfır ve sapmaların karelerinin toplamı minimumdur.

Örneğin, aşağıdaki sınıflanmış serinin aritmetik ortalaması,

$$\bar{x}_i = \frac{\sum x_i f_i}{\sum f_i} = \frac{43}{10} = 4,3$$

x_i	f_i	$x_i f_i$
2	2	4
4	3	12
5	4	20
7	1	7
	10	43

Başlık eklemek için tıklatın

Kişi	Oran	Sabit Tutulan	Kullanılacak Ortalama
Tanık	Yol / Zaman	Zaman	Aritmetik Ortalama
Emin	Yol / Zaman	Yol	Harmonik Ortalama
Hasan	Maliyet / Miktar	Miktar	Aritmetik Ortalama
Selen	Maliyet / Miktar	Maliyet	Harmonik Ortalama

Geometrik ortalama

Duyarsız ortalama