



DERS ADI: Biyoistatistik

ÜNİTE ADI: Temel İstatistik ve Biyoistatistik Kavra

ÜNİTE NO: 1

YAZAR: Prof. Dr. ÖMER AKBULUT

İstatistik Nedir?

İstatistik kelimesi üç farklı anlamda kullanılmaktadır. Bunlar;

- 1) En genel anlamda; üretim, tüketim, nüfus, sağlık, eğitim, tarım, trafik vb. konularda, olguları gösteren veri topluluklarına ve bu verilerin özetlenmesi ve değerlendirilmesi işlemlerine istatistik denir.
- 2) Bilimsel anlamda; araştırmalardan elde edilen verilerin değerlendirilmesi, özetlenmesi, sunumu, bazı analiz ve tahminlerin yapılması ile ilgili prensipleri açıklayan metot bilimine istatistik denir.
- 3) Örnek (örneklem) verilerden hesaplanan ortalama (\bar{x}), varyans (S^2) vb. ölçülere de istatistik denir.

İstatistik biliminin temel kavramları nelerdir?

Populasyon (Kitle, Anakütle): Üzerinde durulan karakterle ilgili tüm elemanları içine alan elemanlar topluluğuna populasyon denir.

Parametre: Populasyonun elemanları üzerinden hesaplanan ortalama (μ), varyans (σ^2), regresyon katsayısı (β) gibi ölçülere parametre denir.

Örnek (Örneklem): Populasyondan şansa bağlı olarak çekilen ve populasyonu temsil etme nitelik ve niceliğindeki elemanlar topluluğuna örnek veya örneklem denir.

Birim: Olayın izlendiği veya incelendiği en küçük populasyon ögesine birim denir. Fert, hasta, deney hayvanı, kan örneği, öğrenci, birim örnekleridir.

Veri: Gözlem, sayım ve ölçüm sonucu elde edilen rakam, harf, sembol, işaret vb. değerlere veri denir.

Değişken: Gözlem, sayım, ölçüm ve değerlendirme sonucunda elde edilen verilerin atandığı çokluklara değişken denir.

İstatistiğin ham maddesi verilerdir. Her değişken veridir, ancak her veri değişken değildir.

Değişkenler Latin alfabesinin X, Y, U, Z gibi son ve büyük harfler ile sembolize edilirler.

Değişkenler kesikli, sürekli, bağımlı, bağımsız, ara ve kontrol değişkeni olarak sınıflandırılmaktadır.

Kesikli Değişken: Ölçülen karakter ölçüm ölçeği üzerinde sadece belli noktalara atanabiliyorsa bu tip değişkenlere kesikli değişken denir. Kesikli değişken bir değer veya sayıdır. Örnek: Bayan-Erkek; 1, 2, 3 vb.

Sürekli Değişken: Ölçülen karakter sayı doğrusu üzerinde belirli bir aralıktaki bütün noktalara atanabiliyorsa bu tip değişkenlere sürekli değişken denir. Sürekli değişken bir aralıktır. Örnek: Kan pH değeri 7.20 - 7.45 arasında, bebek doğum ağırlığı 1.8- 6.5 kg arasında değerlere atanır. Bağımsız değişken sebep, bağımlı değişken sonuç değişkenidir. Örneğin; Vücut ağırlığı sebep, yüksek kolesterol sonuç değişkenidir.

Ölçme ve Ölçek Tipleri

Ölçme: Ele alınan değişkenin, gözlem, sayım, tartım vb. sonuçlarının sembollerle özellikle sayısal değerler ile ifade edilmesi işlemine ölçme denir.

Ölçek tipleri

Adlandırma Ölçeği (Nominal scala): Değişkenlere ad verilerek, yapılan ölçmedir. Adlandırma harf, kelime rakam veya sembol kullanılarak yapılabilir. En zayıf ölçektir. Bu ölçek ile belirlenen değişkenler, aynı zamanda kesikli değişkenlerdir. Bu ölçeğe sınıflı ölçek de denir.

Dereceleme Ölçeği (Ordinal scala): Değişkenlere ad verilmenin yanında değişken ölçüm değerlerinin sıralanabildiği ölçmedir. Adlandırma ölçeğinden daha gelişmiş bir ölçektir. Bu ölçeğe sıralı ölçek de denir. Dereceleme ölçeği ile ölçülen özelliklerin değişken tipide kesikli değişkendir.

Aralık Ölçeği (Interval scala): Skalar veya metrik bir ölçektir. Ölçeğin oluşturulmasında ölçek, gerçek bir noktadan değil, kabul edilen bir noktadan başlatılarak oluşturulur. Bu ölçek adlandırma

ve dereceleme ölçeğinden daha gelişmiş bir ölçektir.

Oran Ölçeği (Ratio scala): Tüm metrik ölçekler (gr, cm vb) oran ölçeği grubunda yer alır. Bu ölçek sıfır (0) başlangıç noktasından başlatılarak oluşturulur. Diğer üç ölçeğe göre en gelişmiş ölçek tipidir. Aralık ve oran ölçekleri sürekli değişkenleri ölçmede kullanılır.

Araştırma, İstatistik ve Bilgi İlişkisi

Araştırma yapmanın amacı yeni bilgilere ulaşmaktır. Bu süreçte elde edilen veriler günümüzde istatistik metotlarla değerlendirilmektedir. İstatistik yöntemler uygulanmayan araştırma sonuçları tartışmalıdır ve günümüz bilim anlayışında kabul görmez.

Olgular hakkında karar verme ve yeni bilgilere ulaşma sürecinde istatistiğin kullanılmasının dört temel amacı vardır. Bunlar;

Durum tespiti yapmak,

Karşılaştırma yapmak,

Tahmin etmek,

İlişki kurmaktır.

İstatistiğin Sınıflandırılması

İstatistik; tanımlayıcı istatistik ve çözümleyici istatistik olarak iki ana gruba ayrılır.

İstatistik bütün bilimlerin ortak veri değerlendirme metodudur. İstatistik; fen, sağlık ve sosyal bilimlerin tüm alanlarında geniş uygulama alanı bulmuştur. Her bir alan istatistik ile eşleştirilerek ilgili alana özel istatistik dalları geliştirilmiştir. Örnek olarak,

Ekonomiye uygulanan istatistik → Ekonometri

Biyolojiye uygulanan istatistik → Biyometri veya Biyoistatistik

Sosyolojiye uygulanan istatistik → Sosyometri

olarak adlandırılmaktadır.

Biyoistatistik

Biyoistatistik, istatistik bilim dalının biyolojik değişkenlere özellikle sağlık alanına uygulanmasıdır.

Biyoloji ve özellikle sağlık alanındaki olaylar ile ilgili verilerin toplanması, bu verilerin özetlenmesi ve yorumlanması, bu alanlardaki değişkenler için test ve tahminlerin yapılması ve nedensellik ilişkilerinin aranmasını inceleyen istatistiğe Biyoistatistik denir. Bu bağlamda Biyoistatistik, sağlık hizmetlerinin planlanmasında, koruyucu hizmetlerin, tanı ve tedavi işlemlerinin etkinliğinin artırılmasında, bu alanlarla ilgili bilimsel çalışma tekniklerinin düzenlenmesinde biyoistatistik yaygın kullanım alanına sahiptir.

Veri kaynakları:

İstatistik veya biyoistatistik çalışmalarda kullanılan veriler elde edilmiş metotlarına göre;

Kayıt verileri

Sayım verileri,

Anket verileri ve

Planlı araştırma (deneysel ve yarı deneysel) verileri olarak sınıflandırılmaktadır.

Bu veriler içinde en güvenilir veriler, planlı araştırma verileridir.

Türetilmiş Veriler

Oran, hız yüzde, binde on binde gibi veriler ve artış oranı (hızı) azalış oranı (hızı) gibi veriler hesaplanarak türetilmiş verilerdir.

Oran: Birimleri aynı iki büyüklüğün birinin diğerine bölümü ile elde edilir. Cinsiyet oranı=Erkek nüfus/Kadın nüfus oranı örneğidir.

Hız: Birimleri farklı iki büyüklüğün birinin diğerine bölümü ile elde edilir. Cismin hızı=Yol/Zaman veya Fiyat=Mal/Para hız örnekleridir.

Pay ve payda birimleri aynı iki büyüklük zaman sürecinde veya başlangıç değerine göre oranlanıyor ise bu oranlar artış hızı veya azalış hızı olarak ta nitelendirilebilirler. Nüfus artış hızı bu duruma örnektir.

Verilerde Aranılan Özellikler

Veriler en az maliyetle elde edilmelidir.

Veriler doğru, güvenilir, gerekli duyarlılıkta ve hassasiyette ölçülmüş veya belirlenmiş olmalıdır.

Veriler kısa sürede elde edilmelidir.

Rakamların (sayıların) yuvarlaklaştırılması

Ondalıkla ifade edilen sayıların belirli seviyede ondalık basamaklarına kısıtlama uygulanması işlemi yuvarlaklaştırma olarak adlandırılır.

Yuvarlaklaştırma işlemi;

Yuvarlaklaştırma

Aşağı yuvarlaklaştırma ve

Yukarı yuvarlaklaştırma

olmak üzere üç farklı şekilde yapılmaktadır. Burada sadece yuvarlaklaştırma anlatılacaktır.

Yuvarlaklaştırma işleminde önce rakamların hangi basamakta yuvarlaklaştırılacağına karar verilir.

İkinci aşamada yuvarlaklaştırma yapılacak ondalıktan sonraki basamaktaki rakam 5 ve 5'ten büyük ise (5,6,7,8,9) karar verilen basamağa 1 eklenir. Bu sayı 5'ten küçük ise (0,1,2,3,4) karar verilen basamak olduğu gibi bırakılır.

Yuvarlaklaştırma işleminde veri veya sonuç değerinden bilgi kaybının oluşmaması esastır. Yani verinin ölçüm veya ifade hassasiyeti kaybolmamalıdır.

Örneğin 7.23678251 olarak ölçülen bir kan pH değeri iki ondalıkla 7.24 olarak yuvarlaklaştırılır.

DERS ADI: Biyoistatistik
ÜNİTE ADI: Tablolar ve Grafikler
ÜNİTE NO: 2
YAZAR: Prof. Dr. ÖMER AKBULUT

Tanımlayıcı istatistiğin amacı; ferler (üniteler) için bilgi veren ancak popülasyon hakkında bilgi vermeyen veya çok sınırlı bilgi veren verileri uygun metotlar kullanarak değerlendirmek, özetlemek, kolay anlaşılır hale getirmek ve yorumlayarak sonuç çıkarabilmektir.

Bu amaçla farklı istatistik metotlar kullanılmaktadır.

Birinci grup tanımlayıcı istatistik metotlar görsel özellik taşıyan istatistik araçlardan oluşmaktadır. Bu ünite de görsel özellik taşıyan istatistik araçları olarak frekans tabloları ve grafikler anlatılmıştır.

Frekans tabloları

Frekans tabloları, verilerin özetlenerek kolay anlaşılır hale getirilmesi fonksiyonuna sahiptir.

Tanımlayıcı istatistikte ve verilerin özetlenmesinde sonuçların sunumunda tablolardan yaygın olarak yararlanılmaktadır.

Frekans tabloları; tek boyutlu tablolar ve iki boyutlu tablolar (çapraz tablolar) olarak iki farklı şekilde oluşturulmaktadır.

Tek Boyutlu Frekans Tabloları

Frekans tabloları satır ve sütunlardan oluşan matris şeklindeki düzenlenmiş yapılardır. Bu tablolarda incelenen değişkene ait sınıflar (alt gruplar, kategoriler) ile bu sınıflara ait frekans, nispi (oransal) frekans ve gerektiğinde eklemeli frekans ve eklemeli nispi frekans ölçüleri yer alır. Tek boyutlu frekans tablosunun ilk sütununda sınıflar (kategoriler) yer alır. İlk satır ise aşağıda tanımları verilen ölçülerin yer aldığı tanımlama satırıdır.

Frekanslar: İncelenen değişkenin ilgili sınıfına veya alt grubuna ait veri sayısı o sınıfın frekans değeridir.

Nispi Frekans, NF: Bir sınıfın frekansının, yani veri sayısının (n), toplam veri sayısına (N) bölünüp, 100 ile çarpılmasıyla elde edilir.

Eklemeli Frekans, EF: Herhangi bir sınıfın frekansının ile o sınıftan önceki sınıfların frekanslarının toplanmasıyla elde edilir.

Eklemeli Nispi Frekans, ENF: Bir sınıfın eklemeli frekansının, toplam veri sayısına (N) bölünüp 100 ile çarpılmasıyla elde edilir.

Sürekli Değişkenlerden Frekans Tablosu Oluşturma

Frekans tablolarında verilerin özetlendiği sınıf değişkeni; kesikli değişken (kan grupları, eğitim düzeyleri vb.) olabildiği gibi sürekli bir değişkenin (kronolojik yaş, kan şekeri değeri, sistolik kan basıncı gibi) belli kurallara bağlı olarak kesikli (kategorik) değişkene dönüştürülmüş formları olabilmektedir. Sürekli değişkenlerin kategorik forma dönüştürülmesi ve frekans tablosu oluşturması işlemi tanımlayıcı istatistikte çok sık uygulama alanı bulmaktadır

Sürekli bir değişken, kategorik bir değişkene dönüştürülerek tek boyutlu frekans tablosu oluşturulurken aşağıdaki işlem sırası takip edilir.

Sınıf sayısı (SS) belirlenir. SS genellikle 5-20 arasında alınır.

Değişim genişliği (DG) hesaplanır. Maksimum değerden, minimum değer çıkartılır $DG = \text{Maksimum Değer} - \text{Minimum Değer}$

Sınıf aralığı (SA) hesaplanır. $SA = DG / SS$.

Sınıf alt limitleri (AL) ve sınıf üst limitleri (ÜL) belirlenir.

Ölçüm hassasiyet birim değerinin yarısı ilgili sınıfın AL değerinden çıkartılarak ve aynı değer aynı sınıfın ÜL değerine eklenerek alt ve üst sınırlar (AS ve ÜS) hesaplanır.

Sınıf orta değerleri (SOD) hesaplanır. Bir sınıfın SOD $AS + ÜS / 2$ veya $AL + ÜL / 2$ formülleri ile hesaplanır.

İlgili sınıfın AL, ÜL değerleri dahil bu aralıktaki veriler sayılarak o sınıfın frekansı belirlenir.

Ölçüm Hassasiyet Birimi (ÖHB): Bir değişkenin verilerinin ölçümünde kullanılan ondalık derecesidir. ÖHB sürekli değişkenin sınıflı forma dönüştürülmesinde dikkate alınması gereken bir bilgidir. Örneğin tam sayı ölçümlerde ÖHB 1'dir. ÖHB tek ondalıkla ölçümlerde genellikle 0.1, iki ondalıkla ölçümlerde 0.01'dir.

Yukarıda anlatılan ve örnekler verilen tek boyutlu frekans tabloları,

Sınıfları niteleyici olan tablolar,

Sınıfları kesikli ancak nicel değerlerden oluşan tablolar,

Sınıf aralıkları eşit olan tablolar,

Sınıf aralıkları eşit olmayan, ancak özel bir anlam ifade eden tablolar ve

Birinci sınıfın alt limiti ve/ya son sınıfın üst limiti olmayan, bir ucu veya iki ucu açık frekans tabloları olarak özetlenebilir.

Aşağıda sürekli değişkenin (BKİ) frekans tablosunda özetlendiği bir tablo örneği verilmiştir. Bu tablo nitelendirici sınıflı, sınıf aralıkları farklı ve iki ucu açık frekans tablolarına örnek oluşturacak özelliktedir.

Örnek Tablo: İki Ucu Açık Sınıflı Frekans Tablosu (Beden Kitle İndeksi Sınıfları, BKİ)

Sınıf	AL - ÜL	Frekans (n)	Nispi Frekans (%)	Eklemleri Nispi F. (%)
Zayıf	< 18.5	8	10.0	10.0
Normal	18.5 - 24.9	47	58.8	68.8
Biraz Kilolu	25.0 - 29.9	15	18.8	87.6
Şişman	30.0 - 39.9	8	10.0	97.6
Obez	40.0 >	2	2.5	100*
		$\Sigma=80$	100*	

*NF toplamı ve ENF toplamı 100 olmalıdır. Ancak sınıflardaki yüzde değerlerde yapılan yuvarlaklaştırmaya bağlı olarak 100.1 veya 99.9 olabilir.

İki Boyutlu Frekans Tabloları (Çapraz Tablolar)

Verilerin özetlenmesinde çok yaygın olarak iki boyutlu tablolardan (çapraz tablolar) yararlanır. İki boyutlu bu tablolara kontenjans tablosu da denilmektedir. Verilerin iki farklı değişkene ait olması durumunda frekanslar iki boyutlu tablolar kullanılarak özetlenebilir.

Çapraz tablolar görünüm olarak çok farklı şekillerde düzenlenebilir.

Çapraz tablolarda değişkenlerden birinin sınıfları yatay (satır) diğerinin sınıfları düşey (sütun) eksene yerleştirilerek matris oluşturulur. Birinci satıra ve birinci sütuna ele alına bu iki değişkenin sınıfları yerleştirilir. Çok kesin bir kural olmamakla birlikte anlatım kolaylığı sağlamak için eğer değişkenlerden biri bağımlı değişken ise, bağımsız değişken yatay eksene (x ekseni) bağımlı değişken düşey eksene (y ekseni) yerleştirilir. Oluşan hücrelerin her birine değişkenlerin ilgili sınıflarının frekansları yazılır. Gerekli görülürse satır toplamları için en sağ tarafa ve sütun toplamları için en alta sütun ve satır toplamları ilave edilebilir.

Ayrıca yine gerekli görülürse her bir hücredeki frekansın altına satır yüzdesi ve sütun yüzdesi ve toplama göre yüzdeler yazılabilir. Yüzde değerlerin hesaplanması tek boyutlu tablolardaki gibidir.

Aşağıda iki boyutlu bir tablo frekans (n) ve yüzdelik değerleri (%) ile sunulmuştur.

Örnek Tablo. Herhangi Bir Lise Grubunda Genel Toplama Göre Başarı veya Başarısızlık % Dağılımı

Başarı Durumu		Mezun Olunan Lise			Toplam
		Fen L.	Anadolu L.	Meslek L.	
Başarılı	n	75	90	80	245
	%	25.0	30.0	26.7	81.7
Başarısız	n	5	10	40	55
	%	1.7	3.3	13.3	18.3
Toplam	N	80	100	120	300
	%	26.7	33.3	40.0	100.0

Tablo yapımında dikkat edilecek hususlar

Tablo başlığı kısa, öz, anlaşılır ve basit olmalıdır.

Tablo başlığına tablonun üstüne yazılmalıdır.

Satır ve sütun değişkenleri ve ölçüleri belirtilmelidir.

Oranlar (% ler) tek başına verilmemeli, gerektiğinde verilmelidir.

N sayısı (fert sayısı) sınırlı ise oran verilmemelidir

Tabloda uygun yerlerde toplamlar(N) verilmelidir.

Grafikler

Verilerin takdim ve özetlenmesinde oldukça yaygın olarak kullanılan bir diğer ana grup araç grafiklerdir. En önemli grafikler,

Histogramlar

Sütun Grafikleri

Çizgi veya Trend Grafikleri

Bölünmüş Daire Grafikleri

Kutu Grafikler ve

Nüfus Piramitleri (Özel Grafikler) olarak sayılabilir.

Grafik çiziminde sıkça bazı hatalar yapılmaktadır. Yapılan en önemli hatalar aşağıda sıralanmıştır.

Bunlar;

Eksenlerin orantısızlığı,

Eksenlerde yer alan değerlerin ve birimlerinin belirtilmemesi,

Şekil veya grafik içindeki detayların açıklanmaması,

4)Şekil ve grafiğin içeriğini belirten başlığın şeklin altına yazılmaması,

Eksenin başlangıç veya sıfır noktasının belirlenmemesi,

Kısaltılan eksenlere kesme “//” işaretinin konmamasıdır.

grafik ile sunulabilmektedir.

DERS ADI: Biyoistatistik

ÜNİTE ADI: Merkezi Eğilim (Yer) ve Değişim (Dağılım) Ölçüleri

ÜNİTE NO: 3

YAZAR: Prof. Dr. CAFER ÇELİK

Verilerin tanımlanmasında kullanılan karakteristik ölçüler bulunmaktadır. Bu ölçüler

Merkezi eğilim (yer) ölçüleri

Değişim (dağılım) ölçüleri

olarak iki ana grupta incelenmektedir.

Yer ölçüleri “Verilerin merkezileştiği noktanın değeri nedir?” sorusuna cevap verirken, değişim ölçüleri “Verilerin değişkenlik durumunun değeri nedir?” sorusuna cevap verir.

Merkezi Eğilim (Yer) Ölçüleri

Merkezi eğilim ölçülerine yer veya yığılma ölçüleri de denilmektedir. Merkezi eğilim (yer) ölçüleri aşağıdaki gibi 3 grupta incelenmektedir. Bunlar;

Ortalamalar

Medyan (ortanca)

Mod (tepe değeri)

Ortalama ölçüsü olarak beş farklı ölçü geliştirilmiştir.

Bunlar;

Aritmetik ortalama

Ağırlıklı ortalama

Geometrik ortalama

Harmonik ortalama ve

Kareli ortalamadır.

Aritmetik ortalama

Dağılımın yerinin belirlenmesinde kullanılır. Tek başına ortalama teriminden aritmetik ortalama anlaşılır.

Aritmetik ortalama (\bar{X}) sembolü ile gösterilir. X_i verilerin değerlerini n , veri sayısını göstermek üzere aritmetik ortalama,

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \text{ formülü ile hesaplanır.}$$

Ağırlıklı (tartılı) ortalama

Verilerin W_i ile ifade edilen bir ağırlığı var ise bu durumda veriler ağırlık değerleri ile ağırlık

değerleri (frekansları) çarpılır ve toplanır. Elde edilen toplam, ağırlıklar toplamına bölünmesi ile ağırlıklı ortalama hesaplanır. Yani X_i : i . verinin gözlenen değeri, W_i : i . verinin gözlenen ağırlığı (frekansı) olmak üzere ağırlıklı ortalama aşağıdaki formül yardımıyla hesaplanır.

$$\bar{X} = \frac{\sum_{i=1}^n W_i X_i}{\sum_{i=1}^n W_i}$$

Geometrik ortalama

Özellikle eşit zaman aralığı ile değişen oranların ortalamasının hesaplanmasında geometrik ortalama kullanılır. Örneğin nüfus artışı, faiz gibi olaylarda ortalama artış hızı geometrik ortalama yöntemiyle hesaplanır.

$X_1, X_2, X_3, \dots, X_n$ gözlenen örnek oranlarını ve n örnek hacmini (veri sayısını) göstermek üzere geometrik ortalama,

$$GO = \sqrt[n]{X_1 * X_2 * X_3 * \dots * X_n}$$

veya

$$GO = \text{Antilog}\left(\frac{1}{n} \sum_{i=1}^n \text{Log}X_i\right)$$

eşitlikleri ile hesaplanır.

Harmonik ortalama

Oran şeklinde türetilmiş verilerde, oran elde edilirken, d/t, eğer pay (d) sabit payda (t) değişken ise oranların ortalaması harmonik ortalama ile hesaplanır. Bunun tersi durumda ise aritmetik ortalama kullanılır. $X_1, X_2, X_3, \dots, X_n$ gözlenen örnek değerlerini ve n örnek hacmini (veri sayısını)

göstermek üzere harmonik ortalama

$$HO = \frac{n}{\frac{1}{X_1} + \frac{1}{X_2} + \dots + \frac{1}{X_n}} = \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{X_i}}$$

formülü yardımıyla hesaplanır.

Üç ortalama arasında **AO ≥ GO ≥ HO** ilişkisi vardır.

Kareli ortalama

Kareli ortalama, veri grubundaki bütün değerlerin karelerinin aritmetik ortalamasının karekökü alınarak hesaplanır. $X_1, X_2, X_3, \dots, X_n$ gözlenen örnek değerlerini ve n örnek hacmini (veri sayısını) göstermek üzere kareli ortalama

$$KO = \sqrt{\frac{X_1^2 + X_2^2 + X_3^2 + \dots + X_n^2}{n}} = \sqrt{\frac{\sum_{i=1}^n X_i^2}{n}}$$

formülü ile hesaplanır.

Medyan (ortanca)

Verilerin küçükten büyüğe veya büyükten küçüğe sıralanması halinde ortaya düşen değer medyandır. Veri sayısı veri sayısı çift ise ortaya düşen iki değer ortalaması medyan (Md) olarak adlandırılır. n gözlenen veri sayısını göstermek üzere;

n tek ise → $\frac{n+1}{2}$ inci değer medyandır.

n çift ise → $\frac{n}{2}$ inci değer ile $\frac{n}{2} + 1$ inci değer ortalaması medyandır.

Medyan ortalamaya nazaran uç değerlerden etkilenmez veya çok az etkilenir ve daha az işlem yükü getirir. Sayılan bu yararlarına karşın ortalama gibi analitik bir değere sahip değildir.

Medyanın standart sapması ortalamankinden büyüktür.

Mod (tepe değeri)

Verilerin içinde en çok tekrarlanan (frekansı en büyük olan) değer mod (Mo) olarak adlandırılır.

Mod değerinin anlamlı olabilmesi için gözlem sayısının çok fazla olması gerekir. Bazen veri grubunun birden fazla modu olabilir. Bu durum, ilgili anakütlenin birkaç alt gruptan meydana geldiğini gösterir. Öte yandan, mod veri grubundaki uç değerlerden etkilenmez.

Herhangi bir veri grubuna ait ortalama, mod ve medyan değeri hesaplandıktan sonra veri grubunun simetrik veya çarpık olup olmadığı aşağıdaki gibi belirlenebilir. İncelenen veri grubu;

Simetrik ise → **Mod = Medyan = Ortalama**

Sağa çarpık ise → **Mod < Medyan < Ortalama**

Sola çarpık ise → **Mod > Medyan > Ortalama**

ilişkileri geçerlidir.

Değişim (Dağılım) Ölçüleri

Verilerin yer ölçülerinden uzaklık durumlarını, yani değişkenliklerini belirtmek için diğer bazı ölçülerin kullanılması gerekir. Verilerin değişkenlik durumunu ve dağılım şeklini belirlemek için kullanılan ölçülere dağılım ölçüleri denir.

İki veri grubu ortalamasının eşit olması dağılımlarının aynı olmasını gerektirmez. Buna karar verebilmek için dağılımın merkezi eğilim ölçüsünün yanı sıra değişkenlik ölçüsünün de bilinmesi gerekir.

Değişim Genişliği (Range): Dağılım şeklinin ve değişkenliğinin karşılaştırılması dağılım ölçüleri ile belirlenir. Bu ölçülerden biri değişim genişliğidir.

$$\text{Değişim Genişliği} = \text{En büyük değer} - \text{En küçük değer}$$

Varyans: Verilerin ortalamasından sapmalarının büyüklüğü dağılımın değişkenliğini gösteren iyi bir ölçüdür. Yaygın olarak kullanılan bu değişkenlik ölçüsü varyanstır.

Örnek varyansı (S^2)
$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$

Anakütle varyansı (σ^2)
$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$

Varyansın büyüklüğü veri grubundaki değişkenliğin fazlalığını ve veri grubunun dağılımının yayvanlığını gösterir. Varyans, değişim genişliğinden daha hassas bir ölçüdür.

Öte yandan, varyansın birimi olmadığından bazı durumlarda verilerin elde edildiği birime sahip bir ölçü kullanılması daha uygun olmaktadır. Böyle bir ölçü varyansın karekökü olan standart sapmadır.

Örnek standart sapması (S)
$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}} \Leftrightarrow S = \sqrt{S^2}$

Anakütle standart sapması (σ)
$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}} \Leftrightarrow \sigma = \sqrt{\sigma^2}$

Örnek standart sapmasının ve varyansının paydasında bulunan n-1 ifadesi serbestlik derecesi olarak adlandırılır. Varyans ve standart sapma bütün veriler kullanılarak hesaplandığından uç değerlerden etkilenir.

Verilerin ortalamasına nispeten değişkenliğinin büyüklüğünü ifade etmek için değişim katsayısı ölçüsü kullanılır. Değişim katsayısı, standart sapmanın ortalamaya oranıdır.

Örnek değişim katsayısı

$$DK = \frac{S}{\bar{X}} * 100$$

Anakütle deęişim katsayısı

$$DK = \frac{\sigma}{\mu} * 100$$

Ortalamaları birbirinden farklı olan anakütlelerin deęişkenliklerinin karşılaştırılmasında deęişim katsayısı ölçüsü kullanılmalıdır. Deęişim katsayısının büyüklüğü arttıkça istenilen deęerden uzaklaştığı söylenebilir.

Bir olayın açıklanmasında hangi merkezi eğilim ve deęişim ölçüsünün kullanılacağı deęişkenin dağılımının normal dağılıma yakın olup olmamasıyla da ilgilidir. Bununla birlikte, veri grubunun ortalaması kullanıldığında ortalama ile birlikte standart sapmasının $(\bar{X} \pm S)$, medyanı kullanıldığında ise en küçük deęer ile en büyük deęerin (medyan; en küçük deęer – en büyük deęer) verilmesi genel bir kural olarak kabul edilmektedir. Böylece açıklamada verilen bilgilerin daha doğru anlaşılabilmesi, yani karşı tarafın yanlış yönlendirmemesi sağlanabilir.

DERS ADI: Biyoistatistik

ÜNİTE ADI: İhtimal ve İhtimal Hesaplamaları

ÜNİTE NO: 4

YAZAR: Prof. Dr. ERKAN OKTAY

İhtimal (olasılık) kavramı hayatımızın her alanında karşımıza çıkabilecek olaylar için kullanılır. Bir olayın olması mümkün olduğu gibi olmaması da mümkün ise bu olayın ihtimale konu olan bir olaydır. N fert (birim) ihtiva eden bir populasyon içinde belli bir X özelliğini taşıyan n tane birim varsa, bu populasyondan rasgele bir birim alındığında bu birimin X özelliğini taşıması ihtimali,

$$P(X) = \frac{n}{N}$$

şeklinde hesaplanır. Bir X değişkeninin nispi frekansı ile bu olayın ortaya çıkma ihtimali arasında yakın bir ilişki vardır.

Bir olayın meydana gelmesi ihtimali 0 ile 1 arasında değişir. İhtimalin sıfır olması, söz konusu olayın meydana gelmesinin mümkün olmadığını, 1 olması ise olayın kesinlikle (yani %100) meydana geleceğini ifade eder. 0 ve 1 durumlarında ihtimalden bahsedilemez. 0'a yakın ihtimal zayıf ihtimal ve 1'e yakın ihtimal ise kuvvetli ihtimaldir.

Bir olayın mümkün bütün hallerinin ihtimalleri toplamı 1'e eşittir. Mesela, bir olayın ancak A, B, C ve D gibi dört yolla meydana gelebilmesi ve bu yollara ait ihtimallerin de sırayla P_A , P_B , P_C ve P_D olması halinde,

$$P_A + P_B + P_C + P_D = 1$$

olur. Yine, X olayının meydana gelme ihtimalini p, meydana gelmeme ihtimalini,

$$1 - p = q$$

şeklinde tarif edersek,

$$p + q = 1$$

olur. Bu tür ihtimallere birbirini tamamlayan ihtimaller denir.

Bu ünite de ihtimaller konusu Biyoistatistik dersinin gerektirdiği kapsamda ve oldukça özet anlatılacaktır.

İhtimallerle ilgili bazı tanımlar

Basit ve Bileşik İhtimaller

Tek bir olayın sonuçları ile ilgili ihtimaller basit ihtimallerdir. Mesela, yarın yağmur yağması ihtimali, bir sınıftan tesadüfi olarak seçilen bir öğrencinin gözlüklü olması ihtimali gibi ihtimaller ayrı ayrı düşünüldüğünde basit birer ihtimaldir. İki veya daha fazla olayın birlikte vuku bulması ihtimali ise bileşik ihtimaldir. Aynı şekilde ikiden fazla olaydan bazılarının bazıları ile birlikte vuku bulması ihtimali de bileşik ihtimaldir. Bileşik ihtimal hesaplarına konu olan olaylar;

Bir arada meydana gelebilen olaylar,

Birbirini engelleyen olaylar,

olarak iki gruba ayrılır.

Örnek Uzayı

İstatistiki bir olayın mümkün olan bütün sonuçlarının oluşturduğu sete örnek uzayı denir ve S ile gösterilir. Örnek uzayındaki her bir sonuç, söz konusu örnek uzayının bir elemanıdır. Örnek uzayı sınırlı sayıda elemana sahipse, karışmamaları için elemanlar birbirlerinden virgülle ayrılıp parantez içerisinde gösterilebilir. Madeni bir para atıldığında mümkün iki sonuçla karşılaşacağımız için örnek uzayı,

$$S = \{Y, T\}$$

şeklinde yazılabilir.

Kontenjans Tabloları ve Venn Diyagramları

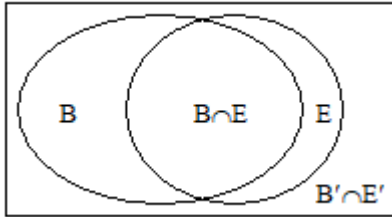
Örnek uzayını başlıca iki yolla gösterebilir. İncelenecek olayları çapraz sınıflandırma yoluyla kontenjans tablolarında gösterilebilir. Mesela, memurlar arasında kredi kartı kullanımını yaygınlaştırmaya çalışan kredi kartı şirketleri bir yılsonunda memurlar arasından tesadüfi olarak 200'ünü seçerek bunlara banka kredi kartı ve/veya seyahat ve eğlence kredi kartı kullanıp

kullanmadıklarını sormuş olsunklar. Alınan cevaplar aşğıdaki kontenjans tablosunda gösterilebilir.

	Seyahat ve Eğlence Kredi Kartı	
Banka Kredi Kartı	Evet	Hayır
Evet	60	60
Hayır	15	65

Olayların birlikte ve ayrı ayrı vuku bulma durumlarını Venn Diyagramı denilen iç içe daireler yardımıyla da gösterilebilir.

Yalnız banka kredi kartına sahip olanlar, B; yalnızca seyahat ve eğlence kredi kartına sahip olanlar, E; hem banka ve hem de seyahat kredi kartına sahip olanlar, $B \cap E$; ; hem banka ve hem de seyahat kredi kartı olmayanlar ise $B' \cap E'$ ile gösterilmek üzere Venn Diyagramı aşğıdaki gibi çizilebilir.



B olayının gerçekleşme ihtimali $P(B)$; E olayının gerçekleşme ihtimali $P(E)$; B ve E olaylarının ikisinin birden gerçekleşmesi ihtimali, $P(B \cap E)$; ; B ve E olaylarının ikisinin birden gerçekleşmemesi ihtimali, $P(B' \cap E')$; B veya E olaylarından birinin gerçekleşmesi ihtimali, $P(B \cup E)$ 'dir.

İhtimal Hesaplama Kuralları

İhtimal hesaplamaları, temel bazı kurallar kullanılarak yapılır.

Toplama Kuralı

X_1, X_2, \dots, X_n birbirini engelleyen n tane olay ve bu olayların meydana gelme ihtimalleri de sırayla P_1, P_2, \dots, P_n olmak üzere, bu olaylardan birinin veya diğerinin meydana gelmesi ihtimali, $P_1 + P_2 + \dots + P_n$

olur. Bu kurala toplama kuralı denir. Birlikte vuku bulmaları mümkün olmayan A ve B gibi iki olaydan birinin veya diğerinin vuku bulması ihtimali,

$$P(A \cup B) = P(A) + P(B)$$

şeklindedir.

A ve B olayları bazı hallerde birlikte de meydana gelebilir. Bu durumda, bu olaylardan birinin veya diğerinin meydana gelmesi ihtimali,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

şeklinde hesaplanır. Yani, olayların birlikte meydana gelme ihtimali, olayların ayrı ayrı meydana gelme ihtimallerinin toplamından çıkarılır.

Çarpma Kuralı

Bir olayın vuku bulması bir başka olayın gerçekleşme şansına bağıl değilse, bu gibi olaylara bağımsız olaylar denir. X_1, X_2, \dots, X_n gibi n tane bağımsız olayın ihtimallerini P_1, P_2, \dots, P_n ile gösterirsek, bu n olayın birlikte meydana gelme ihtimali,

$$(P_1) \cdot (P_2) \cdot \dots \cdot (P_n)$$

olur. Bu kurala çarpma kuralı denir. A ve B gibi iki bağımsız olayın birlikte vuku bulması ihtimali,

$$P(A \cap B) = P(A) \cdot P(B)$$

şeklinde ifade edilir.

Şartlı İhtimal ve Bayes Teoremi

Birlikte meydana gelme özelliğindeki bazı olaylar birbirinden tamamen bağımsız olmayabilir.

Eğer bir olayın meydana gelme ihtimali, başka bir olayın meydana gelme durumuna bağımlı ise, bu durumda şartlı ihtimal kuralı uygulanır. A olayı meydana geldiğinde B'nin ihtimali olarak ifade edilen şartlı ihtimal $P(B/A)$ şeklinde gösterilir.

Bu ihtimal Bayes yaklaşımı ile $P(B/A) = [P(B) \cdot P(A/B)] / P(A)$ formülü ile hesaplanır.

İhtimal Dağılım Tablosu

Bir X olayının meydana gelmesinde mümkün olan haller; X_1, X_2, \dots, X_n ve bu hallerin meydana gelme ihtimalleri de sırayla; P_1, P_2, \dots, P_n ise, söz konusu olaya ait ihtimal dağılım tablosu aşağıdaki gibi olur.

X_i	X_1	X_2	X_n	Toplam
$P(X_i)$	P_1	P_2	P_n	1

İstatistik Olayın N Sayısını Belirleme Metotları

Permütasyon ve Kombinasyon

Bazı durumlarda ihtimal değerini hesaplamak için olayın meydana geldiği yol sayısını (N) ve istenen olayın vuku bulabileceği yol sayısını (k) tespit etmek gerekir. Bunun için permütasyon, ve kombinasyon işlemlerinden yararlanır.

Permütasyon

(n) bireyden her seferinde (r) adedi alınarak oluşturulacak serilerde sıralama önemli ise permütasyon kullanılır. Buna göre incelenen n bireyden her seferinde r adedi alınarak kaç farklı dizinin oluşturulacağı

$${}^n P_r = \frac{n!}{(n-r)!}$$

şeklinde hesaplanır.

Kombinasyon

(k) bireyden her seferinde sıra gözetilmemek kaydıyla (r) tanesi alınarak yapılacak serilerin sayısının belirlenmesine de kombinasyon kullanılır. Burada incelenen n bireyden her seferinde r adedi alınarak kaç farklı dizinin oluşacağı;

$${}^n C_r = \frac{n!}{(n-r)!r!}$$

formülü ile hesaplanır.

DERS ADI: Biyoistatistik

ÜNİTE ADI: Kesikli Populasyon Dağılımları

ÜNİTE NO: 5

YAZAR: Prof. Dr. ERKAN OKTAY

Kesikli değişkenler sayılar ile sembolize edildikleri zamanı belirli değerler arasında sadece tamsayı değerler alabilen değişkendir. Kesikli değişkenler kesikli popülasyon dağılımlarını meydana getirir.

Bir ailedeki çocuk sayısı 0 olabilir, 1 olabilir ve fakat 1.5 olamaz. Çocuk sayısı belirli tamsayı değerlerle ifade edilebilir. Bu örneğimizdeki değişkenimizin alabileceği muhtemel sonuçlar 0, 1, 2, 3, 4, ve 5+'tir (5+= Beş ve daha fazla çocuklu aile).

Değişkenin sonuçlarından birinin gözlenmesi şansa bağlı olarak ortaya çıkar. Bu sebeple bu değişkene, tesadüfi kesikli değişken de denir.

X değişkeni tesadüfi (rasgele, rassal) bir değişken olmak üzere, X değişkenine ait mümkün sonuçları ve bu sonuçların meydana gelme ihtimallerini gösteren dağılımlara kesikli ihtimal dağılımları denir.

Kesikli değişkene ait ihtimaller dağılımının ortalaması da μ ile gösterilir. Ortalamaya beklenen değer de denir ve $E(X)$ ile gösterilir. Kesikli tesadüfi değişkenin ortalaması, kesikli değişken değerleri ile bu değişken değerlerinin ihtimallerinin çarpılıp toplanması sonucu elde edilir. Kesikli tesadüfi değişkenin ortalaması,

$$\mu = E(X_i) = \sum_{i=0}^n X_i P(X_i)$$

formülü yardımıyla elde edilir.

Kesikli tesadüfi değişkenin standart sapması, kesikli ihtimal dağılımının değişkenliğinin bir ölçüsüdür ve s sembolü ile gösterilir. Standart sapmanın büyük olması kesikli değişken değerlerinin daha geniş bir aralıkta dağılım gösterdiğini ifade eder.

Kesikli tesadüfi değişkenin standart sapması

$$\sigma = \sqrt{\sum_{i=0}^n X_i^2 P(X_i) - \mu^2}$$

veya

$$\sigma = \sqrt{\sum_{i=0}^n (X_i - \mu)^2 P(X_i)}$$

formüllerinden biri ile hesaplanabilir.

Binom Dağılımı

Bir X olayının meydana gelmesinde sadece iki hal söz konusu ise bu olayın binom dağılımı gösterdiği söylenir. Uygulamada başarılı veya başarısız, kusurlu veya kusursuz, yazı veya tura, erkek veya kız gibi iki sonuçlu olaylar binom dağılımı gösterirler. Binom dağılımında başarılı olma ihtimali, p; başarısızlık ihtimali ise 1-p ile gösterilir.

Bir olayın binom olayı olabilmesi için p ihtimalinin olaydan olaya değişmemesi gerekir.

Üzerinde durulan olayın n denemede x defa meydana gelmesi ihtimali,

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, 2, \dots, n$$

eşitliğiyle hesaplanır. Formülde

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

şeklinde açılır. Burada x değerleri 0, 1, 2, ..., n gibi kesikli değerler alabileceğinden ve sadece bu değerler için nokta ihtimalleri hesaplanabileceğinden binom dağılımı kesikli bir ihtimal dağılımıdır.

Binom Dağılımının Karakteristikleri

Binom dağılımının parametresi p'dir. Bu dağılımın ortalaması, varyansı, standart sapması, aşağıdaki eşitliklerle hesaplanır:

$$\text{Ortalama } (\mu) = np$$

$$\text{Varyans } (\sigma^2) = np(1-p)$$

$$\text{Standart sapma } (\sigma) = \sqrt{np(1-p)}$$

Poisson Dağılımı

Poisson dağılımı da binom dağılımı gibi kesikli bir ihtimal dağılımıdır. Bu dağılımın bir diğer ismi Küçük İhtimaller Dağılımı'dır. Poisson olayları binom olaylarına çok benzer. Binom dağılımından farklı olarak bu dağılımda üzerinde durulan olayın meydana gelme ihtimali çok küçüktür. Belirli bir alanda, belirli bir uzunlukta, belirli bir hacimde veya belirli bir büyüklük (sayı) içinde nadir olarak ortaya çıkan olaylar Poisson popülasyonuna uygunluk gösterir. Aşağıda bazı tipik poisson popülasyonu örnekleri verilmiştir.

Bir kitapta hatalı problem çözümü,

Bir konserve fabrikasında kusurlu şişe üretimi,

Bir diş kliniğinde hatalı diş çekilme durumu,

Poisson dağılımına konu olan olaylar binom popülasyonunda olduğu gibi iki sonuçludur. Ancak olayın tekrar sayısı, n'in büyümesi p'nin de küçülmesi halinde, binom formülü yerine poisson formülü kullanılır. Daha net bir ifadeyle, np değeri 5'ten küçük olduğunda binom dağılımı poisson dağılımına dönüşür. Belirli bir poisson olayının ihtimali,

$$P(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad x = 0, 1, 2, \dots, n$$

formülü ile hesaplanır. Burada λ , Poisson dağılımının ortalamasıdır. Bu dağılımın varyansı da λ 'ya eşittir. Poisson dağılımının ortalaması $\lambda = np$ formülü yardımıyla hesaplanır. Poisson formülündeki e sabiti, yaklaşık olarak 2.71828 değerine eşittir.

Uygulamada yukarıdaki formülün sonuçlarını doğrudan veren tablolar geliştirilmiştir. Bu tablolardan biri kitabın arkasındaki ekler bölümünde verilmiştir.

Poisson Dağılımının Karakteristikleri:

Poisson dağılımının parametresi λ 'dır.

$$\text{Ortalama } (m) = \lambda = n.p$$

$$\text{Varyans } (s^2) = \lambda$$

Poisson dağılımının ortalaması (μ) , varyansına (σ^2) , eşit olan bir dağılımdır.

$$\text{Standart sapma } (\sigma) = \sqrt{\lambda}$$

Poisson dağılımı normalden daha diktir ve sağa çarpıktır. λ büyüdükçe dağılım simetriğe yaklaşır.

Diğer Kesikli Popülasyon Dağılımları

Ünitenin başında ifade edildiği gibi çok sayıda teorik kesikli popülasyon dağılımı geliştirilmiştir.

Aşağıda diğer bazı kesikli popülasyon dağılımları sıralanmıştır. Bu dağılımların açıklamaları müfredat dışı bırakılmıştır. Bu dağılımların bazıları,

Bernoulli Dağılımı

Çok Terimli Dağılım

Geometrik Dağılım

Negatif Binom Dağılımı
Hipergeometrik Dağılım
Düzgün (Uniform) Dağılım olarak sayılabilir.

DERS ADI: Biyoistatistik
ÜNİTE ADI: Sürekli Populasyon Dağılımları
ÜNİTE NO: 6
YAZAR: Prof. Dr. ÖMER AKBULUT

Sürekli değişkenlerde sürekli populasyon dağılımlarını oluşturur. Biyoistatistiğe konu olan olaylarda binlerce sürekli değişken mevcuttur. Her değişken kendine özgü bir dağılım gösterir. İstatistik analizde değişkenler bu değişkenlerin uyum gösterdikleri teorik populasyon dağılımları ile incelenir. Çok sayıda teorik sürekli populasyon dağılımı geliştirilmiştir. Bu derste sürekli populasyon dağılımlarının önemlileri anlatılmıştır.

Sürekli Populasyon Dağılımları

Sürekli değişkenler, sürekli populasyon dağılımları ile incelenir. Diğer bir ifade ile sürekli değişkenlerden sürekli populasyon dağılımları geliştirilmiştir.

En yaygın sürekli populasyon dağılımı normal dağılımdır. Bu dağılım teorisi detaylı olarak anlatılacaktır. Diğer önemli bazı sürekli populasyon dağılımları olarak;

Log normal dağılım,

Gamma dağılım,

Beta dağılımı sayılabilir.

Normal Dağılım

Dağılımın adı, dağılım fonksiyonunu geliştiren araştırmacıların ismine atfen Gauss Dağılımı veya Gauss-Laplace Dağılımı olarak bilinir. Dağılımın şekli düzgün simetrik bir yapı gösterir ve sürekli dağılımlarının en önemlisidir. Bu dağılıma ait fonksiyonun şekli, çanın kesitine benzediği için, dağılımın diğer bir adı Çan Eğrisi olarak da adlandırılır.

Genellikle, gr, cm, cm³, kg m² vb. ağırlık, uzunluk, hacim, alan ve ünite gibi metrik olarak belirlenen özellikler normal dağılım gösterirler. Daha genel bir ifade ile aralık ve oran ölçeği ile belirlenen veriler normal dağılıma uygunluk gösterir. Dağılım bu özelliği nedeniyle geniş bir kullanım alanına sahiptir.

Dağılımın ihtimal fonksiyonu,

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

dir. Formülde

e = 2.718, ve $\pi = 3.1416$, sabitleri olmak üzere,
x = Ele alınan ve normal dağılım gösteren değişkeni,
 μ = Değişkeninin ortalamasını,

σ^2 = Söz konusu değişkenin varyansını

σ = Değişkenin standart sapmasını göstermektedir.

Normal dağılım fonksiyonunun şeklini ve yerini değişkenin μ ve σ^2 parametreleri belirlemektedir.

Standart Normal Dağılım (z Dağılımı)

Normal dağılım fonksiyonundaki $\frac{x-\mu}{\sigma}$ ifadesini z'ye eşitleysek $z = \frac{x-\mu}{\sigma}$ şeklinde yazarsak, normal dağılım fonksiyonundan,

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

şeklinde ifade edilen standart normal dağılım fonksiyonu elde edilir. Çünkü bu

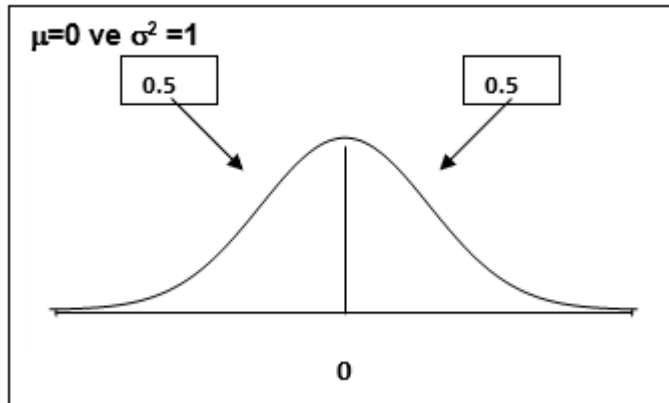
$$\sigma_z^2$$

fonksiyonun değişken parametresi z'nin ortalaması $\mu_z = 0$ ve varyansı $\sigma_z^2 = 1$ dir.

Dağılım simetrik olup, dağılımın tepe noktasının yatay ekseni kestiği nokta 0 noktası olup eğrinin

altındaki alan 1'e eşittir. (Aşağıdaki şekli inceleyiniz).

Dağılım simetrik olduğu için dik çizginin sağ ve sol tarafındaki alanlar birbirine eşittir. Eğrinin altındaki standart alan bir olup, sağ taraf 0.5 sol taraf 0.5 birimdir. Eğrinin altındaki alan 100 kabul edilirse, dik çizginin sağ tarafı %50, sol tarafı %50 dir. Diğer bir ifade ile Standart z dağılımında $P(z > z_0) = P(z < -z_0) = 0.5$ dir. Bu dağılım için $P(z > z_1)$ ihtimalleri 0- sonsuz (∞) aralığı için istatistik dokümanlarda Z tablosu adı altında sunulur.



Şekil 6.2. Standart Normal Dağılım

Standart z değerleri kural olarak iki ondalıkla hesaplanmalıdır. Standart Normal Dağılım (z Dağılımı) aşağıdaki durumlarda yaygın olarak kullanılır. Bunlar;

Normal dağılım ile ilgili problemlerin çözümünde kullanılır.

Normal dağılıma uymayan veya dağılımı bilinmeyen populasyonlardan çekilen örneklerin ortalaması ile ilgili problemlerin çözümünde kullanılır.

Binom, poisson vb. dağılımların normale yaklaşım gösterdiği durumlarda bu dağılımlarla ilgili problemlerin çözümünde kullanılır.

Merkezi Limit Teoremi

Dağılımı bilinmeyen veya normal dağılıma uymayan populasyonlardan çekilen (n) fertlik örneklerin ortalamalarının ortalaması populasyon ortalamasına, örnek ortalamalarının varyansı ise populasyon varyansının örnekteki fert sayısına bölümüne eşittir. Merkezi limit teoremine göre bu örnek ortalamalarının dağılımıda yaklaşık normal dağılım göstermektedir. Bu durumda z dağılımına göre,

$$z = \frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \text{ şeklinde ifade edilir.}$$

Ayrıca yine merkezi limit teoremine göre, bir özellik çok sayıda faktörün bir sonucu olarak şekilleniyor ise bu özelliğin dağılımı normal dağılım gösterir. Örnekleme gerekirse insanlarda biyolojik bir değişken (özellik) olan boy uzunluğunun dağılımı normal dağılıma uyar. Çünkü boy uzunluğunun şekillenmesine genetik faktörlerin yanında beslenme şekilleri, spor yapma veya yapılan spor çeşidi, yaşanan coğrafya, gibi çok sayıda faktör etkilidir. Bu faktörlerin hepsinin olumlu etkisine maruz kalan kişi uzun boylu, olumsuz etkisine maruz kalan kişi ise kısa boylu olur. Halbuki fertler daha çok bu faktörlerin bir kısmının olumlu, diğer bir kısmı olumsuz etkisine maruz kalacaktır. Nadir olarak tüm olumlu faktörler veya olumsuz faktörler bir fertte bir araya gelecektir. Bu nedenle toplumda insanlar nadiren çok kısa boylu veya çok uzun boylu olurken daha çok belirli bir boy etrafında (Örneğin 170 cm) yığılım gösterir ve boy özelliğinin dağılımı normal dağılım gösterir.

Bu nedenle teorem esas alınarak çok sayıda faktörün etkisi sonucu şekillenen ve metrik ölçümlerle belirlenen boy, kilo, kolesterol, glikoz vb. değişkenler merkezi limit teoremi esas alınarak normal dağılım yaklaşımı ile analiz edilmektedir.

Normal Dağılımla İlgili Diğer Dağılımlar

Normal dağılımdan türetilmiş standart normal z dağılımından başka t dağılımı, (Student t dağılımı)

χ^2 dağılımı ve

F dağılımı

gibi dağılımlarda mevcuttur.

Student t Dağılımı

Bu dağılım, örnek hacminin küçük olduğu durumlar için ($n < 30$) Gosset (1908) tarafından geliştirilmiştir. Standart normal dağılımdaki populasyon varyansı yerine örnek verilerden hesaplanan örnek varyansı, S^2 , esas alınarak, küçük örnekler için geliştirilmiştir. Dağılımın

ortalaması, populasyon ortalaması μ 'ye eşittir. Ortalamanın standart sapması olarak adlandırılan $s_{\bar{x}} = \sqrt{S^2 / n}$ şeklinde tahmin edilerek örnek ortalamalarının populasyon ortalamasından farkı $(\bar{x} - \mu)$, dağılımın standart hatası cinsinden

$$t = \frac{\bar{x} - \mu}{\sqrt{S^2 / n}} = \frac{\bar{x} - \mu}{s_{\bar{x}}} \text{ formülü ile elde edilir.}$$

Bu şekilde hesaplanan t değerleri de $-\infty$ ile $+\infty$ arasında değişir ve simetriktir. t dağılımı standart z dağılımına göre biraz basıktır. Ancak t değerleri genelde -4 ile 4 arasında değişir ve t'lerin ortalaması 0 olup, normal dağılım gösterir. Serbestlik derecesine (sd) bağlı olarak farklı ihtimal değerleri için dağılımın kritik t değerleri t tablolarında sunulur. t dağılımı, z dağılımına göre daha basık olup, örnek büyüklüğüne bağlı olarak, serbestlik derecesi büyüdükçe z dağılımına yaklaşır. Sonsuz serbestlik derecesinde t ve z kritik tablo değerleri eşittir.

χ^2 , Dağılımı (Ki-kare Dağılımı)

Standart normal dağılımdaki z_i lerin ($z_1, z_2, z_3, \dots, z_k$) karelerinin toplamı k sd'li χ^2 dağılımını

oluşturur. χ^2 dağılımının ortalaması, $E(\chi^2) = k$ ve varyansı $\text{Var}(\chi^2) = 2k$ dir. Dağılım pozitif ve sağa çarpıktır Eğrinin şekli k'ya bağlıdır ve k büyüdükçe dağılım simetrikleşir. Serbestlik

derecesine göre Ki-kare değerleri χ^2 tablolarında verilir.

F Dağılımı

İki ki-kare değişkeninin veya iki varyansın oranı ile elde edilen bir dağılımdır. Dağılım pozitif ve hafif sağa çarpıktır.

Bu dağılımın ait pay ve payda serbestlik derecesine bağlı kritik F tablosunda sunulmuştur. Pay ve paydanın serbestlik derecesi büyüdükçe dağılım simetrikleşir.

DERS ADI: Biyoistatistik

ÜNİTE ADI: Örneklem ve Örneklem Metotları

ÜNİTE NO: 7

YAZAR: Prof. Dr. ÖMER AKBULUT

Örneklem Nedir?

İstatistiksel değerlendirmeler için araştırmacı nitelik ve nicelik bakımından popülasyonu temsil eden bir alt grup oluşturur. Popülasyonu temsil eden bu alt gruba örnek adı verilir. Populasyondan alt grubun (örneğin) oluşturulması işlemine ise örneklem denir. Araştırmalarda genellikle örnekten elde edilen veriler kullanılarak test ve tahminler yapılır, popülasyon hakkında karar verilir. Test ve tahminlerin doğru olması için popülasyonla örnek arasındaki ilişkiyi çok iyi kurmak gerekir.

Örneklem Metotları

Popülasyonun sınırlı, sınırsız, homojen veya heterojen olması durumuna göre örneklem metotları geliştirilmiştir. Örneklem amaçlı örneklem (bilinçli) ve şansa bağlı örneklem (rassal) olarak iki temel esasa göre yapılır.

Amaçlı Örneklem Metotları

Amaçlı örneklem metotları yargısal örneklem metotları olarak ta isimlendirilirler. Amaçlı örneklemede belirli ön yargılara göre, yine belirli özellikleri taşıyan bireyler örneğe alınır.

Amaçlı örneklem metotları daha çok sosyal bilimlerde kullanılmaktadır. Bu metotlar;

Monografik örneklem,

Kota örneklem,

Yoğunluk örneklem,

Güdümlü örneklem,

Kartopu örneklem metotları olarak adlandırılırlar.

Şansa Bağlı Örneklem Metotları

Şansa bağlı örneklemin esası; popülasyondaki bütün fertlerin örneğe girme şansının eşit olmasıdır. Ayrıca fertler popülasyondan tamamen şansa bağlı olarak seçilir. Analitik istatistikte tahmin ve analiz metotları örneklemin şansa bağlı olması esasına dayanır. Bu şansa bağlı örneklem metotları daha yaygın kullanım alanına sahiptir. Bu gruptaki başlıca örneklem metotları aşağıda kısaca anlatılmıştır. Bunlar;

Basit Şans Örneklemesi

Sistemik Örneklem

Tabakalı Şans Örneklemesi

Küme Örneklemesi

Çok Basamaklı Örneklem

Kombinasyon Örneklem metotlarıdır.

Basit şans örnekleme

Popülasyon homojen ise (sınırlı veya sınırsız olabilir) basit şans örnekleme uygulanır. Bu işlem için popülasyondaki fertleri temsil eden sayılar veya isimler bir torbaya konur ve örnek sayısı kadar iadeli çekiliş yapılır. Eğer popülasyon sınırsız ise (çok büyük) iadesiz çekiliş yapılır.

Sistemik örneklem

Popülasyon sınırlı ve fertler tabi olarak sıralı ise sistemik örneklem yapılır. Sistemik örneklemede popülasyon homojendir ve her ferdin bir sıra veya seri numarası vardır. Popülasyon büyüklüğü (N) belirlidir. Popülasyon büyüklüğü, örnek büyüklüğüne (n) bölünür. Elde edilen değer tam sayıya yuvarlatılarak sıralı verilerin dilim büyüklüğü (k) belirlenir. Sıralı veriler n kadar k dilimine ($k_1, k_2, k_3, \dots, k_n$) ayrılmış olur. İlk k_1 dilimi içinden yine şansa bağlı bir sayı (c) belirlenir. Bu c sıra değerine sahip fert ilk dilimi örnekte temsil eder. Sonraki dilimlerde $k+c, 2k+c, 3k+c, 4k+c, \dots$ ve $nk+c$ sıra numaralı fertler sistemik olarak örneğe alınır. Bu şekilde yapılan örneklem sistemik örneklemedir.

Tabakalı şans örnekleme

Popülasyon heterojen olması durumunda önce popülasyonda homojen alt gruplar (tabakalar) oluşturulur. Bu tabakalardan büyüklükleri (n_j) ve heterojenlikleri ölçüsünde örneğe fert şansa bağlı olarak seçilerek örnek oluşturulur.

Küme örnekleme

Popülasyonun oldukça büyük ve fertlerin rasgele kümeler oluşturduğu durumlarda bu örneklem

metodu kullanılır. Bu metotta küme tümüyle veya belirli oranda fert örneğe ayrılır.

Çok basamaklı örnekleme

Populasyon basamaklar halinde giderek küçülen alt gruplardan (katman-basamak) oluşması durumunda bu örnekleme kullanılır. Ferdin içinde bulunduğu en alt grup küme örneklemede olduğu gibi topluca örneğe alınır.

Kombinasyon örnekleme

Örnekleme hatasını en aza indirmek amacıyla yukarıda anlatılan şansa bağlı örnekleme metodlarının iki veya daha fazlasının birlikte uygulanması şeklinde yapılan örnekleme metodudur.

Örnek Büyüklüğünün Tahmini

Araştırmacıların en çok karşılaştıkları problemlerden biri yürütülecek araştırmada kullanılacak denek, (birim, fert, ünite) sayısı ne olmalıdır? İstatistiksel test ve tahmin metodlarının en önemli şartlarından biri örneğin yeterli büyüklükte veriden (fertten) oluşturulmasıdır. Araştırmacılar hipotezlerine doğru cevap verecek büyüklükte örnek kullanmaları gerekir. Yetersiz örnekleme hatalı bulgulara neden olurken gereğinden fazla büyük örnek daha fazla maliyet, daha fazla işgücü ve emek gerektirir.

Ayrıca deney hayvanları ile yapılan çalışmalarda kullanılacak hayvan sayısı, istatistik metodlarla tahmin edilen gerekli örnek sayısından fazla olmamasına özen gösterilir. İstatistik metodlarla belirlenen sayıdan fazla deney hayvanı kullanımı hayvan hakları bakımından etik dışı olarak kabul edilmektedir.

Bütün bu nedenlerle örnek büyüklüğünü etkileyen faktörlerin bilinmesi ve örnek büyüklüğünün doğru tahmini önemlidir.

Örnek Büyüklüğünü Etkileyen Faktörler

Örnek büyüklüğü ve örnek hacmi eş anlamlı kavramlardır. Örnek büyüklüğü belirlenirken özellikle üç husus dikkate alınır..

Değişkendeki varyasyon (σ^2 veya s^2)

Kabul edilen bilimsel hata düzeyi (α)

Tahminin sapması veya güvenirliliği ($d = \bar{X} - \mu$)

İncelenecek populasyonda bir örneklik arttıkça varyans küçülür. Varyansı küçük olan özellikler daha küçük örnekle temsil ettirebilirler.

Bilimsel çalışmalarda hata düzeyi (α) genellikle 0.05 olarak alınır. Alfa değeri $\alpha = 0.05$ değerinden $\alpha = 0.001$ değerine doğru küçüldükçe örneğe alınacak fert sayısı artar.

Tahminin sapması ile populasyon parametresi arasındaki sapma etki büyüklüğü (effect size) olup d ile gösterilir. d değeri büyütüldükçe örnek büyüklüğü (fert, birim sayısı) azalır.

Ortalama İçin Örnek Büyüklüğünün Tahmini

Sürekli değişkenler için örnek büyüklüğünün tahmini popuasyon varyansının s^2 bilinmesi durumu çok nadir olduğundan örnek büyüklüğü, S^2 kullanılarak hesaplanmaktadır. Hesaplama,

$$n = \left(\frac{t^2 S^2}{d^2} \right) \text{formülü kullanılır.}$$

Formülde kullanılan t^2 açık yazılımı, $t^2_{\alpha/2(n-1)}$ t tablosunda $\alpha/2$ ihtimal düzeyinde ve n-1 serbestlik derecesindeki cetvel değerinin karesidir. S^2 örnek varyansı, d ise yine tahmin edilecek populasyon ortalaması ile hesaplanan örnek ortalaması arasında izin verilecek sapma miktarıdır.

Örnek varyansı S^2 de bilinmiyorsa ön bir çalışma ile oluşturulan örnek veriler kullanılarak örnek varyansı S^2 hesaplanarak, formüldeki yerine yazılarak hesaplama yapılır.

Populasyon Oranı İçin Örnek Büyüklüğünün Belirlenmesi

Oran için örnek büyüklüğünün belirlenmesinde N bilinmiyorsa birinci N biliniyorsa aşağıdaki ikinci formül kullanılır.

$$n = \frac{\hat{p}\hat{q}(z_{\alpha/2}^2)}{d^2} \quad n = \left(\frac{N \hat{p} \hat{q} (z_{\alpha/2}^2)}{(N-1) d^2 + \hat{p} \hat{q} (z_{\alpha/2}^2)} \right)$$

Formüllerde p olayın teorik görülme sıklığı (oranı) veya ön çalışma ile örnek veriler üzerinden hesaplanan oran değeri, $q=1-p$, $z^2_{\alpha/2}$ z tablosunda $\alpha/2$ ihtimal düzeyindeki tablo değerinin karesidir. Formüldeki d ise populasyon oranı ile tahmin edilecek oran arasında kabul edilen sapma miktarı ve N populasyon büyüklüğüdür.

DERS ADI: Biyoistatistik
ÜNİTE ADI: Nokta ve Aralık Tahminleri
ÜNİTE NO: 8
YAZAR: Prof. Dr. MEHMET TOPAL

İstatistik karar teorisi testler ve tahminler olmak üzere ikiye ayrılır. Tahminler, hakkında hiçbir bilgiye sahip olunmayan populasyon parametresinin değerinin tahmin edilmesini sağlar. Populasyon parametresi hakkında ileri sürülen doğruluğu ve yanlışlığı önceden bilinmeyen varsayımların örnekten elde edilen bilgilerin kullanılmasıyla test edilmesi ise hipotez testleri yardımıyla olur.

İstatistikte tahminler nokta tahmini ve aralık tahmini olmak üzere ikiye ayrılır.

Nokta Tahmini ve Aralık Tahmini

Nokta tahmini, sayı doğrusu üzerinde tek bir noktayı gösterir. Aralık tahmini, sayı doğrusu

üzerinde iki nokta arasında bulunur. Örneğin, bir özelliğin ortalaması $\bar{X}=60$ ise bu değer nokta tahminidir. Bu ortalama %90 güvenle 50 ile 70 arasında olacağı bulunmuşsa, bu aralık tahminidir. Aralık tahmini bulunurken her zaman nokta tahminleri kullanılır. Nokta tahminleri bir diğer ifadeyle istatistik değerler populasyon parametrelerinin tahminleridir. Aşağıda nokta tahminleri ve tahmin edilen parametre örnekleri verilmiştir.

Ölçü	Tahminci (İstatistik)	Tahmin edilen (Parametre)
Ortalama	\bar{X}	μ
Varyans	S^2	σ^2
Standart sapma	s	σ
Oran	\hat{p}	P veya π
Korelasyon katsayısı	r	ρ
Regresyon katsayısı	b	β
...

Nokta Tahmininin Özellikleri

Nokta tahminlerinin belirli özellikleri vardır. Bu özellikler aşağıda kısaca özetlenmiştir.

Sapmasızlık: Bir istatistiğin beklenen değeri, parametresine eşitse bu istatistik veya nokta tahmini

sapmasızdır. Aritmetik ortalamasının beklenen değeri $E(\bar{X})$ populasyon ortalaması μ 'ye eşit olduğundan aritmetik ortalama sapmasız bir tahmindir.

Kararlılık: Örnekteki fert sayısının artmasıyla nokta tahmini, populasyon parametresine yaklaşıyorsa ilgili tahmin kararlıdır. Örnek büyüklüğü arttıkça, aritmetik ortalama populasyon parametresine daha yaklaştığı için, aritmetik ortalama mod ve medyana göre daha kararlıdır.

Etkinlik: Varyansı en küçük nokta tahmini en etkin tahmindir. Aritmetik ortalamasının varyansı

$\sigma_{\bar{X}}^2 = \sigma^2/n$, medyanın varyansından $V_{med} = \sigma^2\pi/2n$ daha küçük olduğundan aritmetik ortalama medyana göre daha etkin tahmindir.

Yeterlilik: Nokta tahmininin örnekteki bilgileri kullanma özelliğidir. Örnekteki bilgileri en fazla kullanan nokta tahmini en yeterlidir.

Aralık Tahmini (Güven Sınırları)

Aralık tahmini (güven sınırları); nokta tahmini kullanarak, populasyon parametresinin belirli bir ihtimalle içinde bulunduracağı sınırların tahmini şeklinde tanımlanır. Populasyon parametresine ait güven sınırları genel olarak

$a < \text{parametre} < b$

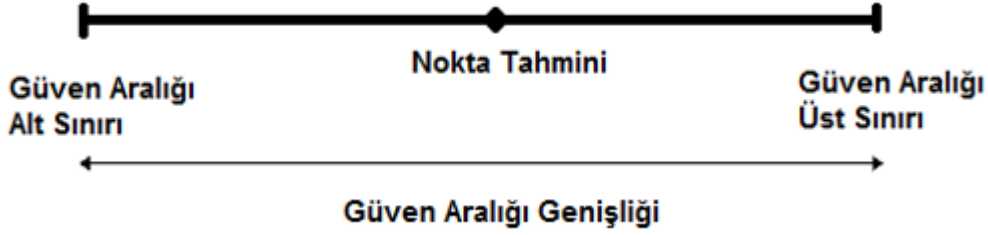
şeklinde gösterilir. Belirli bir α ihtimali ile bu eşitlik

$P(a < \text{parametre} < b) = 1-\alpha$

şeklinde tanımlanabilir.

Eşitlikteki a değeri ilgilenilen parametrenin alt sınırını, b değeri üst sınırını ve $1-\alpha$ değeri ise güven katsayısını göstermektedir. Güven katsayısı, güven aralığının populasyon parametresini içermeye ihtimalini, α değeri ise populasyon parametresinin sınırlar dışında kalma ihtimalini göstermektedir. Aralık tahmininde üst sınır ile alt sınır arasındaki fark (b-a) ise güven aralığı olarak ifade edilir.

Nokta ve aralık tahmini için aşağıda verilen şekil izlenmelidir.



Aralık tahmininin genel formülü şu şekilde ifade edilir.

Aralık Tahmini = Nokta Tahmini \pm Kritik Değer \times Standart Hata

Bir Ortalamanın (μ) Güven Sınırlarının Tahmini

Bir ortalamanın güven sınırlarını tahmin ederken iki durum söz konusudur. Eğer populasyona ait varyans değeri (σ^2) biliniyorsa Z dağılımı, bilinmiyor ise t dağılımı kullanılır. Bir ortalamanın güven sınırlarını tahmin edebilmek için;

Varyans (σ^2 veya S^2) bilinmeli

Örnek verilere ait ortalama değer hesaplanmalı

Örnek büyüklüğü (n) bilinmelidir.

Z Dağılımı Kullanılarak Bir Ortalamanın (μ) Güven Sınırlarının Tahmini

Dağılımı bilinmeyen populasyonlardan elde edilen örneklere Z dağılımı uygulanırken Merkezi Limit Teoremi kullanılır. Ortalaması μ ve varyansı σ^2 olan herhangi bir populasyondan şansa bağlı olarak çekilen n birimlik örneklerin ortalamalarının dağılımı normal, ortalaması μ ve varyansı $\sigma_{\bar{x}}^2 = \sigma^2/n$ dir.

Buradan $\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}$ ve Z eşitliği $Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$ şeklinde olur.

Belirli bir hata (α) veya güvenle ($1-\alpha$) Z değerinin içinde bulunabileceği sınırlar için Z değeri

$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1-\alpha$ eşitliğinde yazılarak gerekli işlemler yapılarak ortalama için alt ve üst güven sınırları için

$$\mu_{ÜS} = \bar{x} + Z_{\alpha/2} * \sigma_{\bar{x}}$$

$$\mu_{AS} = \bar{x} - Z_{\alpha/2} * \sigma_{\bar{x}}$$

formülleri elde edilir.

Ortalama için güven sınırlarının tahmininde bu basit eşitlikler kullanılır.

Bir değişken (özellik) için %95 güven aralığı %90'a göre daha geniştir. Yani güven düzeyi ($1-\alpha$) büyüdükçe dolayısıyla hata (α) küçüldükçe güven aralığı genişlemektedir.

t Dağılımı Kullanılarak Bir Ortalamanın (μ) Güven Sınırlarının Tahmini

Populasyon varyansı bilinmediğinde örnek varyansı S^2 hesaplanabildiği ve örnek büyüklüğü $n < 30$ olduğu durumlarda, ortalamanın güven sınırlarının tahmininde t dağılımı kullanılır. Dağılımın genel formülü;

$$t = \frac{\bar{x} - \mu}{S_{\bar{x}}} \quad \text{ve} \quad S_{\bar{x}} = \sqrt{\frac{S^2}{n}}$$

Belirli bir ($1-\alpha$) güven veya hata (α) seviyesinde ve n-1 serbestlik derecesinde t dağılımına göre bir ortalamanın güven sınırları aşağıdaki eşitliklere göre hesaplanır,

$$\mu_{ÜS} = \bar{x} + t_{\alpha/2, n-1} * S_{\bar{x}}$$

$$\mu_{AS} = \bar{x} - t_{\alpha/2, n-1} * S_{\bar{x}}$$

Bir Oranın Güven Sınırlarının Tahmini

Oranların aralık tahmininde örnek oranının (p) populasyon oranını temsil edebilmesi için örneğin yeterli büyüklükte ($n \geq 30$) olması gerekir.

($1-\alpha$) güven veya α hata seviyesinde populasyon oranın güven sınırları hesaplanırken;

$$p_{\text{ÜS}} = \hat{p} + Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$p_{\text{AS}} = \hat{p} - Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

eşitlikleri kullanılır.

Varyansın Güven Sınırlarının Tahmini

Varyansın güven sınırlarının tahmininde Ki-Kare dağılımı kullanılır.

Dağılımı normal, ortalaması μ varyansı σ^2 olan bir popülasyondan şansa bağlı çekilen bir örneğin varyansı S^2 kullanılarak popülasyon varyansının $1-\alpha$ güven düzeyinde bulunabileceği sınırlar tahmin edilir.

Popülasyon varyansının üst güven sınırı ($\sigma_{\text{ÜS}}^2$) ve alt güven sınırı (σ_{AS}^2);

$$\sigma_{\text{ÜS}}^2 = \frac{(n-1)S^2}{\chi_{1-\alpha/2; n-1}^2}$$

$$\sigma_{\text{AS}}^2 = \frac{(n-1)S^2}{\chi_{\alpha/2; n-1}^2}$$

formülleri kullanılarak hesaplanır.

DERS ADI: Biyoistatistik
ÜNİTE ADI: Temel Parametrik Testler
ÜNİTE NO: 9
YAZAR: Prof. Dr. ÖMER AKBULUT

İstatistik karar teorisinin en önemli konularından biri hipotez testleridir. Hipotez testleri; ortalamaların karşılaştırılması, oran testleri, ilişki testleri, uyum testleri, homojenlik testleri gibi çok farklı amaçlar ve şekillerde yapılmaktadır.

Ayrıca hipotez testleri parametrik testler nonparametrik testler (parametresiz testler) ve Ki-kare testleri olmak üzere üç grup altında sınıflandırılarak yapılmaktadır.

Bu ünite de hipotez ve hipotez testi kavramı anlatılacaktır. Hipotezlerin kurulması, testin yönünün belirlenmesi ilkeleri açıklanacak ve temel parametrik testler ile hipotezlerin nasıl test edileceği anlatılmıştır.

Hipotez Kavramı Ve Hipotez Testi

İstatistik anlamda “Hipotez”; araştırmacının öngördüğü iddiadır. Araştırmanın başlangıcında doğruluğu veya yanlışlığı kestirilemeyen ön görülere veya iddialara hipotez denir.

Hipotez Testi Nedir?

Hipotezin (iddianın, varsayımın) örnekten elde edilen veriler kullanılarak, belirli bir hata payı ile doğrulanmasına hipotez testi denir. Genel bir ifade ile hipotez testi yapmak araştırma yapmaktır. Bu nedenle önce analize tabi tutulacak verileri elde etme tekniklerini yani araştırma tekniklerini bilmek gerekir.

Temel (Ho) ve Karşıt (H1) Hipotezler

Bir araştırmada, Temel Hipotez (Ho) ve Karşıt Hipotez (H1) olmak üzere iki tip hipotez kurulur. Mevcut veya bilinen durumun ifade edilmesi ile Ho hipotezi oluşturulur. Ho hipotezi fark yoktur, eşittir benzerdir, ilişki yoktur ifadeleri kullanılarak genelde negatif anlamlı kurulur. İstatistik yazılımda eşittir(=) sembolü kullanılır.

Bazı temel hipotez örnekleri şu şekildedir.

Yüksek rakımın kan basıncına etkisi yoktur (Ho: $\mu = \mu_0$).

Sigara içmenin kalp spazmına etkisi yoktur (Ho: $P = P_0$).

Kolesterol ile yaş arasında ilişki yoktur. (Ho: $r = \mu_0$)

H1 hipotezi araştırmacının iddiası veya öngörüsüdür. Araştırmacı iddiasını istatistiksel notasyonla ifade ederek H1 hipotezini oluşturulur. Bu aşamada Ho hipotezindeki fark yoktur, eşittir, benzerdir (genel anlamı ile (=) eşitlik) ifadelerine karşılık,

Fark vardır, eşit değildir, değişmiştir, benzer değildir ilişki vardır veya

Artmıştır, yükselmiştir, büyümüştür, çoğalmıştır vb. veya

Azalmıştır, düşmüştür, küçülmüştür, eksilmiştir vb. ifadeler kullanılarak H1 hipotezi kurulur.

Yukarıdaki ifadeler dikkat edilirse a) grubu ifadeler eşitsizlik (\neq) b) grubu ifadeler büyüktür ($>$) ve c) grubu ifadeler ($<$) sembollerine karşılık gelmektedir. İstatistik terminolojisinde eşitsizlik (\neq) ifade eden H1 hipotezleri çift yönlü büyüktür veya küçüktür ($>$ veya $<$) ifade eden H1 hipotezleri tek yönlü hipotezler olarak nitelendirilir.

Bazı karşıt hipotez örnekleri yukarıdaki HO hipotezleri için şu şekildedir.

Yüksek rakım kan basıncını değiştirir (H1: $\mu \neq \mu_0$; iki yönlü).

Sigara içme kalp spazmını artırır (H1 $P > P_0$; tek yönlü).

Kolesterol ile yaş arasında pozitif bir ilişki vardır. (H1: $r > 0$)

Hipotez Testlerinde Hata Düzeyi

Bilimsel çalışmalarda hipotezler test edilirken müsaade edilen en yüksek hata düzeyi %5 olup, bu hata düzeyi alfa işareti (α) ile gösterilir. Bilimsel hata düzeyleri, bu hatalar için kullanılan asteriksler (yıldız işaretleri) ve anlamları aşağıdaki gibidir.

$\alpha = 0.05$	* (önemli),
$\alpha = 0.01$	** (çok önemli),
$\alpha = 0.001$	*** (çok çok önemli),
$\alpha = 0.10$	+ (marjinal önemli).

Bir hipotez testi yürütülürken iki tip hata yapma ihtimali her zaman söz konusudur. Bu hatalar I. Tip hata ve II. Tip hata olarak adlandırılır.

I. Tip Hata: Populasyonda H_0 hipotezi geçerli iken test sonucunda H_1 hipotezinin kabul edilmesi ile yapılan hataya I. Tip hata denir. I. Tip hatanın seviyesi α 'dır.

II. Tip Hata: Populasyonda H_1 hipotezi geçerli iken test sonucunda H_0 hipotezinin kabul edilmesi ile yapılan hataya II. Tip hata denir. İkinci tip hata β ile sembolize edilir ve müsaade edilen en yüksek β hatası % 20 (0.20) dir.

I. tip hata azaldıkça II. tip hata artar. Aynı şekilde tersi de doğrudur. Bu iki tip hatayı birlikte kontrol etmek için alınması gereken en önemli üç tedbir vardır. Bunlar,

Araştırma homojen materyal üzerinde yapılmalıdır.

Örnekleme doğru yapılmalı ve örnek hacmi yeterli olmalıdır.

Veri analizinde güçlü test seçilmelidir.

Test Sonucu	Populasyondaki Durum	
	H_0 Doğru	H_1 Doğru
Test Sonucu H_0 Kabul	Doğru Karar	II. tip hata (β hatası)
Test Sonucu H_1 Kabul	I. tip hata (α hatası)	Doğru Karar

Hipotez Testlerinin Temel Aşamaları

Hipotez testi yapılırken, hipotezlerin kurulması, bilimsel hata düzeyinin belirlenmesinin yanında verilerin uyum gösterdiği teorik populasyon dağılımının da belirlenmesi gerekir. Çünkü test işlemleri bu dağılımın teorik esaslarına göre yürütülür. Hipotez testlerinde yapılan işlemler ve işlem sırası aşağıdaki gibi özetlenebilir.

1. Aşama: Hipotezler kurulur (H_0 ve H_1).
2. Aşama: Dağılım tespit edilir ve kritik cetvel değeri belirlenir.(CD)
3. Aşama: Test istatistiği hesaplanır.(HD)
4. Aşama: Test istatistiği (HD) ve kritik cetvel değeri (CD) karşılaştırılarak karar verilir.

Temel Parametrik Testler

İstatistiksel veri analizinde en yaygın kullanılan parametrik testler karşılaştırma testleridir. Bu testler Standart normal z dağılımı ve Student t dağılımı kullanılarak yapılır. Oranlar ve ortalamaların karşılaştırılmasında yaygın kullanılan parametrik testler aşağıda sınıflandırılmıştır. Oranların karşılaştırılmasında kullanılan en yaygın hipotez testleri;

Bir oranın hipotez testi (p) ve

İki oran farkının hipotez testi (p_1-p_2) dir.

Oranların hesaplandığı ve analitik olarak değerlendirildiği çalışmalarda, örnek hacmi büyük olduğu için, bu grup testlerde z dağılımı kullanılır.

Ortalamaların karşılaştırılması söz konusu olan çalışmalarda genellikle populasyon varyansı bilinmediği ve örnek varyansı hesaplanabildiğinden ayrıca örnek hacmi nispeten sınırlı olduğu için Student t dağılımı kullanılmaktadır. t Dağılımı ile yapılan hipotez testleri;

Bir ortalamanın hipotez testi (μ)

iki ortalama farkının hipotez testi ($\mu_1- \mu_2$)

Grup karşılaştırması =Bağımsız iki grup testi

Eşleme testi = Bağımlı iki grup testi

şeklinde sınıflandırılmaktadır.

Burada çok yaygın kullanılan grup karşılaştırması testi özetlenecektir. Diğer testler benzer süreçle gerçekleştirilir.

İki ortalama farkının hipotez testi ($\mu_1- \mu_2$)

İki ortalama farkının hipotez testi iki gruptaki deneklerin birbirinden bağımsız veya bağımlı olmasına göre farklı iki şekilde yapılır. Gruplardaki denekler birbirinden bağımsız ise “Grup Karşılaştırması Testi” uygulanır. Denekler arasında bağımlılık (eşleşme) söz konusu ise “Eşleme testi” uygulanır.

Grup karşılaştırması (Bağımsız iki grup testi)

Bağımsız iki grup ortalaması arasındaki farklılığın testi için gruplar ve gruplardan elde edilen

veriler aşağıdaki özellikleri taşımalıdır. Bu özellikler testin ön şartlarıdır.

Her iki grupta verilerin dağılımı normal dağılıma uygun olmalıdır.

Örnekler yeterli büyüklükte olmalıdır.

Materyal homojen olmalıdır.

Gruplar homojen materyalin benzer iki grubu olmalıdır.

Her iki gruptan hesaplanan varyanslar homojen olmalıdır.

Örnekleme şansa bağlı yapılmalıdır.

Bağımsız iki grubun ortalamalarının karşılaştırılmasında takip edilen teorik işlem sırası daha önce anlatılan hipotez testlerine benzer olarak,

1.Aşama: Hipotezler kurulur.

$$H_0 : \mu_1 = \mu_2 \quad \text{veya} \quad \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 \neq 0, H_1 : \mu_1 - \mu_2 > 0; H_1 : \mu_1 - \mu_2 < 0$$

2.Aşama: Kritik cetvel değeri belirlenir.

$$t_c = t_{\alpha/2(n_1+n_2-2)} \quad \text{iki yönlü test için,}$$

$$t_c = t_{\alpha(n_1+n_2-2)} \quad \text{tek yönlü test için,}$$

3. Aşama: Test istatistiği hesaplanır.

$$t_h = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{S_{\bar{x}_1 - \bar{x}_2}} \quad \text{formüldeki} \quad S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \quad \text{formülü ile hesaplanır.}$$

4.Aşama: Karşılaştırma yapılır ve karar verilir.

$$|t_h| > |t_c| \Rightarrow H_0 \text{ ret } H_1 \text{ kabul edilir.}$$

$$|t_h| < |t_c| \Rightarrow H_0 \text{ kabul edilir.}$$

DERS ADI: Biyoistatistik
ÜNİTE ADI: Ki-kare Testleri
ÜNİTE NO: 10
YAZAR: Prof. Dr. ERKAN OKTAY

Nitel deęişkenlere atanan veriler sayılarak elde edilmiş ise bu deęişkenler arasındaki sebep-sonuç ilişkileri araştırılırken ki-kare testlerine başvurulur.

Örnekleme yoluyla elde edilen rakamların gözlenen deęerlerin teorik (beklenen) deęerlere uygunluk gösterip göstermedięi de ki-kare testi ile tespit edilir. Ki-kare dağılımına dayalı çok sayıda test geliştirilmiştir. Ancak uygulamada yaygın kullanılan başlıca Ki-kare testleri bağımsızlık testi ve homojenlik testidir.

Ki-kare deęişkeni χ^2 sembolü serbestlik derecesi (Sd) veya (v) ile sembolize edilir. Bu ünite de ki-kare serbestlik derecesi için v sembolü kullanılmıştır.

Bir Ki-kare dağılımının ortalaması, dağılımın serbestlik derecesine ve varyansı, serbestlik derecesinin iki katına eşittir. Örneğin, serbestlik derecesi 10 olan bir ki-kare dağılımının ortalaması 10 ve varyansı 20'dir. Ki-kare deęerleri 0 ile $+\infty$ arasında deęişir. Bu sebeple eğri sıfır noktasından başlar ve sonsuza uzanır.

Bağımsızlık Testi

Çok sık karşılaşılan araştırma sorularından biri de iki kesikli (nitel) deęişkenin birbiri ile ilişkili olup olmadığıdır. Örneğin, bir halk saęlığı uzmanı eğitim seviyesinin ile sigara içip içmeme durumları ile veya eğitim seviyesinin gelir dilimleri ile ilişkili olup olmadığını araştırmak isteyebilir. Bu ce benzer ilişkiler Ki-kare bağımsızlık testi ile analiz edilir.

Bir popülasyondaki iki nitel deęişkenin birbirinden bağımsız olup olmadığına karar vermek için şu varsayımların gerçekleşmesi gereklidir. Bunlar;

Veriler, n adet tesadüfi örnekten elde edilmelidir.

Örnekteki gözlemler iki kritere göre çapraz sınıflandırılabilir.

Deęişkenler tabii olarak kategorilere ayrılabilir veya ölçüm deęerleri birbirinden ayrı sayısal kategorilere ayrılabilir.

Bu şartları saęlayan veriler genel olarak bir kontenjans tablosunda yani satır (R) ve sütunlardan (C) oluşan iki boyutlu (RXC) bir tabloda gösterilebilir.

Ki-Kare bağımsızlık testi süreci, hipotez testi sürecinin aynı adımları takip edilerek yürütülür.

Burada tek fark gözlenen deęerler RxC boyutlu tabloda özetlendiğinde bu gözlenen deęerlerin teorik deęerlerinin (beklenen deęerler) hesaplanmasıdır.

Testin aşamaları;

Hipotezler kurulur.

H_0 : İki sınıflama kriteri birbirinden bağımsızdır.

H_1 : İki sınıflama kriteri birbirinden bağımsız değildir.

2) Beklenen deęerler (E_{ij}) hesaplanır

Beklenen deęerler (E_{ij}) basit orantı yaklaşımı ile hesaplanır. Burada n_i i. Satır toplamını, n_j J. Sütun

toplamını, n genel toplamı göstermek üzere, $E_{ij} = \frac{n_i \cdot n_j}{n}$ dir.

3) Test İstatistięi hesaplanır.

Gözlenen deęerler (G_{ij}) ve beklenen deęerler kullanılarak test istatistięi aşağıdaki formül ile hesaplanır.

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \left[\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \right]$$

Test istatistięi, H_0 hipotezi doğru iken, $v = (r - 1)(c - 1)$ serbestlik derecesiyle yaklaşık olarak χ^2 dağılımı gösterir.

Kritik deęer belirlenir

Kritik değer $v = (r - 1)(c - 1)$ serbestlik derecesi ve a (birinci tip hata) önem seviyesine göre Ki-Kare tablosunda (Dağılımın teorik tablosu Ek Tablo 3) yer alan değerdir. Bu değer ki-kare dağılımının sağ ucunda yer alan a kadarlık bölgenin sınırındadır.

5) Karar modeli oluşturulur

Hesaplanan χ^2 test istatistiği $\chi_{\alpha, v}^2$ ile ifade edilen kritik değer karşılaştırılarak karar verilir. Eğer test istatistiği kritik değerden büyük ise H_0 hipotezi ret edilir. Diğer durumda H_0 hipotezini ret edilemez.

2 x 2'lik Kontenjans Tablosu

Her iki sınıflama kriterinden iki kategori bulunduğu, elde edilecek kontenjans tablosunun iki satırı ve iki sütunu olur. Bu tip bir tabloya dört gözlü kontenjans tablosu denildiği gibi 2 x 2'lik kontenjans tablosu da denir. 2x2 boyutlu tablolarda bağımsızlık testi yukarıda anlatılan adımlarla yapıldığı gibi bu test için geliştirilmiş özel bir test istatistiği formülü de vardır.

2 x 2'lik tablolardan χ_h^2 değeri hesaplanırken Yates süreklilik düzeltmesi kullanılır.

Kontenjans Katsayısı

Ki-kare testi ile sadece gerçek ve teorik değerler arasında bir bağımlılığın veya uygunluğun olup olmadığı belirlenir. Fakat, eğer varsa, bu bağımlılığın veya uygunluğun oranını tespit edilemez. Bu oran, kontenjans (uygunluk,bağımlılık) katsayısı ile belirlenir. Kontenjans katsayısı,

$$c = \sqrt{\frac{\chi_h^2}{\chi_h^2 + n}}$$

formülü ile hesaplanır.

Homojenlik Testi

Homojenlik testinde her bir populyasyondan belirli hacimlerde örnekler çekilir ve bu örneklerin belirli bir karakteristiği taşımaları bakımından benzer değişim gösterip göstermediklerine karar verilir. Veriler bağımsızlık testinde anlatılan $R \times C$ tablolarında gösterilir.

Homojenlik testinde ise iki farklı durum söz konusudur.

I.Durum:

Ki-kare homojenlik testinde birinci durumda yapılan işlemler Ki-kare bağımsızlık testinin aynısıdır. Ancak hipotezler ve örnekleme farklıdır. Bağımsızlık testinde başarı durumunun bölümlere bağımlı olup olmadığı test edilirken, homojenlik testinde başarı ve başarısızlık oranları bakımından bölümlere dağılımın homojen olup olmadığı test edilmiştir.

II. Durum

Ki-kare homojenlik testinin ikinci durumunda farklı populyasyon veya alt populyasyonların önceden bilinen veya kestirilen bir orana dağılım bakımından homojen olup olmadıkları kontrol edilir. Burada beklenen değerlerin hesaplanması I. Durum prosedüründen farklılık gösterir.

Test istatistiği alt populasyonlar için (χ_h^2) ve birleştirilmiş örnek için (χ_T^2) ayrı ayrı hesaplandıktan sonra nihai test istatistiği aşağıdaki formül ile hesaplanır.

$$\chi_D^2 = \chi_h^2 - \chi_T^2$$

Diğer Ki-Kare Testleri

Bilimsel çalışmalarda kullanılan diğer bazı ki-kare testleri; Uyum Testi, Fisher'in kesin Ki-Kare testi ve Mc Nemar testleridir.

Uyum Testi

Bilinen veya teorisi olan herhangi bir dağılıma uyum, üniform dağılıma uyum testleri Ki-kare uyum testi ile yapılır. Bu test ayrıca herhangi bir değişkenin Binom, poisson, normal vb. teorik populyasyon dağılımlarına uyumunu kontrol etmek içinde kullanılır.

Fisher'in kesin Ki-Kare testi (Fisher's Exact Test)

2*2 boyutlu(4 hücreli) kontenjans tablolarında beklenen değerlerden biri 5 veya 5 den küçük ise bu veriler Fisher'in Kesin Ki-kare testi ile analiz edilir.

Mc Nemar Testi

Eğer 2*2 boyutlu(4 hücreli) tablolarda sayılarak özetlenen veriler önce-sonra, ilk ölçüm-ikinci ölçüm gibi bağımlılık gösteriyor ise bu veriler Mc Nemar testi ile analiz edilmelidir.

DERS ADI: Biyoistatistik
ÜNİTE ADI: Nonparametrik Testler
ÜNİTE NO: 11
YAZAR: Prof. Dr. ÖMER AKBULUT

Parametrik test şartlarının sağlanmadığı durumlarda şartları daha az olan istatistik yöntemler geliştirilmiştir. Parametresiz testler veya Nonparametrik testler olarak adlandırılan bu testler dağılıştan bağımsız testler olarak ta adlandırılmaktadır.

Nonparametrik testler değişkenin dağılışının normal dağılışa, uygunluk göstermediği, örnek hacminin sınırlı olduğu, varyansların heterojen olduğu durumlarda veri analizinde yaygın olarak kullanılmaktadır. Ayrıca adlandırma ve dereceleme ölçeği ile ölçülen verilerin analizi de nonparametrik testler ile yapılmaktadır.

Nonparametrik testlerin bu avantajlarının yanında en önemli dezavantajı, parametrik testlere göre güçleri zayıftır. II. Tip hata yapma ihtimali nonparametrik testlerde daha yüksektir. Ayrıca bazı kompleks parametrik testlerin alternatifi nonparametrik testler henüz geliştirilmemiştir.

En yaygın kullanılan parametrik testler ve bu testlerin karşılığı (alternatifi) aşağıdaki tablo da verilmiştir.

Önemli Parametrik Testler ve Nonparametrik Test Karşılıkları

Parametrik Test	Nonparametrik Test
Bir Ortalamanın hipotez testi	İşaret testi
Bağımsız İki Ortalamanın hipotez testi	Mann <u>Whitney</u> U testi
Bağımlı İki Ortalamanın hipotez testi	İşaret testi ve <u>Wilcoxon</u> Eşleme testi
Tek yönlü <u>varyans analizi</u>	<u>Kruskal Wallis</u> testi
İki yönlü <u>varyans analzi</u>	<u>Friedman</u> testi

Tek Örnek İçin Parametrik Testler

Parametrik test şartlarının sağlanmadığı durumlarda tek grup verilerin hipotez testi için nonparametrik testlerin kullanımı tercih edilir.

İşaret Testi

İşaret testinde merkezi yer ölçüsü olarak Medyan kullanılır. Bu testte örnek veriler medyandan büyük olanlar (+), küçük olanlar (-) eşit olanlar (0) ile işaretlenir. Veriler +, 0, ve – sembollerine dönüştürülerek işlemler yapıldığı için teste İşaret testi adı verilir.

Hesaplamalarda n örnek büyüklüğünü, k dizide (+) ile işaretlenenlerin sayısını, göstermek üzere, Test istatistiği $n \geq 25$ ve $n < 25$ durumları dikkate alınarak iki farklı şekilde yürütülür.

Kolmogorov –Smirnov Testi (Tek Örnek)

Kolmogorov-Smirnov tek örnek testi, χ^2 uyum testinin alternatifidir. Ki-kare testinin uygulanabilmesi için her bir hücrede beklenen değer frekansının 5'ten az olmaması gerekir. Bunun içinde örneğin yeterli büyüklükte olması gerekir. Halbuki çoğu durumlarda, örnek hacmi yeterli büyüklükte olmayacağı için bu test kullanım alanı bulur.

Testin yapılış aşamaları hipotez testi aşamaları ile aynıdır. F_o : Gözlenen eklemeli nispi frekans, F_e : Beklenen eklemeli nispi frekans olmak üzere, Test istatistiği D ile sembolize edilir ve $D = \max(F_o - F_e)$ dir. Test istatistiği D, Kolmogorov – Smirnov kritik değer tablosu değeri ile (KD) karar verilir. $D > KD$ ise gözlenen frekansların beklenen frekanslara uygun olmadığına karara verilir.

Dizi Sayıları Testi (Run Test)

Tek örnek için yapılan nonparametrik testlerden bir diğeri Dizi Sayıları Testi'dir. Bu test Run test olarak da adlandırılmaktadır. Populasyondan çekilen örneğin şansa bağlı (rasgele, tesadüf) çekilip çekilmediğini test etmek için kullanılan bir testtir.

İki Örnek İçin Nonparametrik Testler

Eşlenik İki Örnek Testi

Karşılaştırma yapılacak eşlenik iki örnek parametrik test varsayımlarının geçerli olmadığı durumlarda nonparametrik testler uygulanır. Parametrik eşleme testinin alternatifidir.

Eşlenik iki örnek durumunda iki farklı test kullanılır. Bunlardan biri Eşlenik Çift İşaret Testi diğeri Wilcoxon Eşlenik Çift İşaret Testidir. Kısaca Wilcoxon Eşleme testi olarak ta adlandırılan bu test daha yaygın kullanılmaktadır Çünkü eşlenik verileri farklarının hem yönünü hem de miktarlarını dikkate alır.

Testin örneklem büyüklüğünü dikkate alan farklı yaklaşımları vardır. Karar modeli standart Z dağılımı veya Wilcoxon değerler tablosu kullanılarak oluşturulur ve karar verilir.

Bağımsız İki Örnek Testi

Bağımsız iki örneğin ortalamasının karşılaştırılmasında, parametrik test varsayımlarının geçerli olmadığı durumlarda nonparametrik teknikler uygulanır. İki bağımsız örnek durumu için kullanılan en yaygın nonparametrik teknikler,

Mann Whitney U Testi

Kolmogrov-Smirnov İki Örnek Testidir.

Mann Whitney U Testi

İki örneğin aynı populasyondan alınıp alınmadığı veya örneklerin alındığı popülasyonların birbirinden farklı olup olmadığını test etmede kullanılır.

Mann Whitney U testinin yapılabilmesi için verilerin en az sıralama ölçeği (ordinal ölçek) ile belirlenmiş olması gerekir. Bu test Küçük Örnek durumu ve Büyük Örnek durumu olmak üzere iki farklı yaklaşımla yapılır.

Eğer iki örnekten büyük örnek 20 veya daha az gözleme sahip ise küçük örnek yaklaşımı izlenir. Eğer iki örnekten birinde veri sayısı 20'den fazla veya her iki örnekte veri sayısı 10'dan fazla ise, büyük örnek yaklaşımı ile veriler analiz edilir. Büyük örnek yaklaşımında karar modelinin oluşturulmasında standart Z dağılımından yararlanır.

Mann Whitney U testi bazı kaynaklarda "U Testi" olarak da adlandırılır.

Kolmogorov - Smirnov İki Örnek Testi

Kolmogorov – Smirnov iki örnek testi, tek örnek testinde olduğu gibi grupların eklemeli dağılımının uygunluğunu inceler. Eğer iki örnek aynı dağılımı gösteren popülasyonlardan alınmışlar veya aynı popülasyonun şansa bağlı iki örneği iseler, örnek veriler popülasyon dağılımından sadece tesadüfi sapmalar gösterecektir. Bu durumda örneklerin eklemeli dağılımları birbirine benzer olacaktır.

Kolmogorov – Smirnov iki örnek testi, iki ayrı durum için farklı şekillerde yapılır

I. Durumda; $n_1, n_2 < 40$ ve $n_1 = n_2$ 'dir. Bu durum küçük örnek durumu olarak adlandırılır. Bu durum için kritik değerler Ek Tablo 9 da verilmiştir.

II.Durumda; $n_1, n_2 > 40$ ve $n_1 = n_2$ eşitliği aranmaz. Bu durum ise büyük örnek durumu olarak adlandırılır.

Bu testin büyük örnek durumu (II. Durum) daha yaygın kullanılmaktadır. Testin test istatistiği ve karar modelinin oluşturulmasında χ^2 dağılımı esas alınır.

İkiden Fazla Örnek Durumu

İkiden fazla bağımsız gruplarda "Kruskal-Wallis Testi", ikiden fazla bağımlı gruplarda "Friedman testi" ile veriler analiz edilir. Bu testlerin teorik aşamaları ve örnek açıklamaları daha ileri biyoistatistik derslerinin konusu olarak düşünülerek bu testlerle ilgili daha ileri anlatım yapılmamıştır.

DERS ADI: Biyoistatistik
ÜNİTE ADI: Regresyon ve Korelasyon Analizi
ÜNİTE NO: 12
YAZAR: Prof. Dr. MEHMET TOPAL

İstatistiksel veri analizinde sürekli değişkenler arasındaki ilişki regresyon ve korelasyon analizleriyle incelenir.

Regresyon Analizi

Regresyon, üzerinde durulan değişkenlerden birinin bağımlı (Y) ve diğerlerinin veya diğerlerinin bağımsız (X_1, X_2, \dots, X_k) olması durumunda Y'nin (X_1, X_2, \dots, X_k)'nin bir fonksiyonu olarak ifade edilmesi ve bu fonksiyona göre X_1, X_2, \dots, X_k 'dan Y'nin tahmin edilmesini sağlar. Dolayısıyla bir fonksiyonel bağıntı olan regresyon; bağımsız değişkenlerdeki değişmelerin bağımlı değişkeni hangi yönde ve ne miktarda etkilediğini belirler.

Bir regresyon denklemi, bir bağımsız değişken içeriyorsa basit regresyon eşitliği, birden fazla bağımsız değişken içeriyorsa çoklu regresyon denklemi olarak adlandırılır.

$$Y = \alpha + \beta X \quad \text{basit doğrusal}$$

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots \quad \text{ise çoklu doğrusal regresyon eşitliğidir.}$$

Basit Doğrusal Regresyon Analizi

Basit doğrusal regresyon modelinde bir bağımlı (Y) ve birde bağımsız (X) değişkeni bulunur. Bağımsız değişken genellikle araştırmacı tarafından kontrol edilebilir fakat bağımlı değişkene araştırmacının hiç bir etkisi olamaz.

Basit doğrusal regresyon modelinde değişkenlerin normal dağılımlı olması, eşit varyanslı olması ve yeterli örneklem gibi varsayımları sağlamalıdır

Regresyon Analizi İçin Parametrelerin Tahmini

En Küçük Kareler Tahmin Metodu

En küçük kareler metoduna göre regresyon parametrelerinin tahmini, doğruyu; noktaların kendisinden ayrılışlarının kareleri toplamını minimum yapacak şekilde tayin etme esasına dayanmaktadır. Populasyonda (X_i, Y_i) gözlemlerine ait doğrusal regresyon modeli aşağıdaki gibi yazılır:

$$Y_i = \alpha + \beta X_i + \epsilon_i$$

Basit doğrusal regresyonda amaç, bağımsız değişken X ile bağımlı değişken Y arasındaki ilişkiyi doğru bir şekilde belirleyecek eşitliği tespit etmektir. Örnekte bağımsız değişken X ile bağımlı değişken Y arasındaki basit regresyon eşitliği,

$$Y = a + bx + e$$

şeklinde yazılır. Eşitlikteki a ve b, sırasıyla α ve β 'nin tahminleridir. Regresyon modelinde,

y_i : Bağımlı değişkeni (Sonuç değişkenini),

x_i : Bağımsız değişkeni (sebebe değişkenini),

a: Regresyon doğrusunun Y eksenini kestiği nokta olup başlangıç değerini,

b: Regresyon katsayısını veya doğrunun eğimini, diğer bir ifade ile x'deki 1 birimlik değişime karşılık y'deki değişim miktarını,

e_i : Şansa bağlı hata değerini göstermektedir.

Örnek için hesaplanan a ve b değerleri α ve β 'nin sapmasız tahminleridir. Örnekler için hesaplanan a değerlerinin ortalaması α 'ya ve b değerlerinin ortalaması β 'ya eşittir
Regresyon denklemindeki a ve b aşağıdaki formüllere göre hesaplanır.

$$b = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

$$a = \bar{y} - b\bar{x}$$

Tahmin edilen a ve b katsayıları $y = a + bx$ denkleminde yerlerine yazılarak doğrusal regresyon

denklemleri kurulmuş olur.

X ve Y değişkenleri için basit doğrusal eşitlik örneğin $Y=11.8 + 0.495X$ şeklinde bulunmuş ise bu eşitlik şu şekilde yorumlanır:

X değişkeni sıfır değerinde iken Y değişkeni 11.8 başlangıç değerine sahiptir. X değişkenindeki bir birimlik artış Y değişkeninde 0.495 birimlik artışa neden olur.

Regresyon eşitliğindeki katsayılar matris sistemiyle de bulunabilir. Bağımsız değişken sayısının iki ve daha fazla olduğu durumlarda a, b1, b2, ... katsayıların tahmini matris yöntemi ile tahmini daha kolaydır.

Regresyon Katsayılarının Önem Testi

Regresyon katsayılarının hipotez testi ve güven sınırlarının tahmininde t dağılımı kullanılır.

Bu katsayıların (a ve b ya karşılık a ve b) istatistiksel anlamlılığı (sıfırdan farklılığı) hipotez testi ile belirlenir. Yapılacak testte hipotez testi süreci aşamaları aynı şekilde dört aşama izlenerek yürütülür.

Tahmin edilen denklemin uyumu diğer bir ifade ile Y değişkenindeki varyasyonun Bağımsız X değişkeni (veya değişkenleri X1, X2, ...) tarafından açıklanma düzeyi “belirleme katsayısı” adı verilen R^2 istatistiği ile belirlenir.

Belirleme katsayısı R^2 0 ile 1 arasında ($0 \leq R^2 \leq 1$) değişir. Belirleme katsayısının 1’e yaklaşması doğrunun uyumunun iyi olduğunu, 0’a yakın olması uyumun olmadığını gösterir.

Korelasyon Analizi

Korelasyon, aralık, oran ve derecelendirme ölçüğü ile belirlenen iki değişken arasındaki ilişkinin ifade edilmesi korelasyon bu ilişkinin ölçüsü ise korelasyon katsayısı olarak adlandırılır. Örnek veriler üzerinden hesaplanan korelasyon katsayısı “r” ve popülasyon verileri üzerinden hesaplanan korelasyon katsayısı “p” ile gösterilmektedir. Korelasyon katsayısı $-1 \leq r \leq +1$ aralığında değer alır. Bu katsayı -1’e yaklaştıkça iki değişken arasında negatif yönde doğrusal bir ilişkinin, +1’e yaklaştıkça iki değişken arasında pozitif yönde doğrusal bir ilişkinin olduğu ve 0’a yaklaştıkça iki değişken arasındaki ilişkinin azaldığı ifade edilir. Korelasyon ölçüsü sebep-sonuç ilişkisi hakkında bilgi vermez.

Korelasyon katsayısının belli aralık değerine göre daha detaylı yorumlanması aşağıdaki gibidir.

r aralığı	Anlamı	r aralığı	Anlamı
0.00 - 0.25	Pozitif, zayıf ilişki	(0.00) - (-0.25)	Negatif, zayıf ilişki
0.26 - 0.50	Pozitif, orta ilişki	(-0.26) - (-0.50)	Negatif, orta ilişki
0.51 - 0.75	Pozitif, iyi ilişki	(-0.51) - (-0.75)	Negatif, iyi ilişki
0.76 - 1.00	Pozitif, yüksek ilişki	(-0.76) - (-1.00)	Negatif, yüksek ilişki

Basit Korelasyon Katsayısının (r) Tahmini,

Normal dağılım gösteren x ve y gibi iki değişken arasındaki Pearson korelasyon katsayısı,

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}} = \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sqrt{\left(\sum X_i^2 - \frac{(\sum X_i)^2}{n}\right) \left(\sum Y_i^2 - \frac{(\sum Y_i)^2}{n}\right)}}$$

formülü ile hesaplanır. Bu formülün uygulanması bilinmelidir.

Korelasyon Katsayısının Önem Testi ve Güven Sınırları

Korelasyon katsayısının hipotez testi ve güven sınırlarının tahmininde t dağılımı kullanılır.

Korelasyon katsayısının anlamlılığı test edilirken yine hipotez testinin dört aşaması uygulanır.

Korelasyon katsayısının güven sınırları

Korelasyon katsayısının üst güven sınırı ($\rho_{\text{üst}}$) ve alt güven sınırı (ρ_{alt}) aşağıdaki eşitlikler kullanılarak tahmin edilir.

$$\rho_{\text{üst}} = r + t_{\alpha/2; (n-2)} \times S_r$$

$$\rho_{\text{alt}} = r - t_{\alpha/2; (n-2)} \times S_r$$

Regresyon ve korelasyon için test ve tahminlerde “serbestlik derecesinin” belirlenmesinde veri

çifti sayısının 2 eksiği alındığına dikkat edilmelidir.

Korelasyon Katsayısı ile Regresyon Katsayısı Arasındaki İlgisi

Regresyon analizinde Y'nin X'e göre regresyon katsayısı b_{YX} ile X'in Y'ye göre regresyon katsayısı b_{XY} farklıdır. Yani, $b_{YX} \neq b_{XY}$ şeklindedir.

Korelasyon analizinde Y'nin X'e göre korelasyon katsayısı r_{YX} ile X'in Y'ye göre korelasyon katsayısı r_{XY} birbirine eşittir. Yani $r_{YX} = r_{XY}$ dir.

Aynı veri setinde karşılıklı regresyon katsayılarının geometrik ortalaması korelasyon katsayısını verir ($r_{YX} = r_{XY} = \sqrt{b_{YX}b_{XY}}$).

DERS ADI: Biyoistatistik
ÜNİTE ADI: Nüfus İstatistikleri
ÜNİTE NO: 13
YAZAR: Prof. Dr. ÖMER AKBULUT

Nüfus; doğum, ölüm, dışa göç ve içe göç olaylarına bağlı olarak şekillenen bir olgudur. Nüfus istatistikleri, sağlık hizmetlerinin etkinliğini belirlemek, denetlemek ve geleceğe ait projeksiyonların yapılabilmesi için özel bir öneme sahiptir.

Oran ve hız, sağlık alanında yararlanılan en önemli göstergelerdir. Bölgenin veya ülkenin sağlık profillerinin belirlenmesinde hız, oran ve diğer istatistiklerden yararlanır Oran (proportion) ve Hız (rate) bir bölgede meydana gelen olayların %, (yüzde) ‰ (binde), ‰0 (on binde) gibi değerlerle ifade edilmesini sağlayan göstergelerdir.

Nüfus Sayımı

Nüfus sayımı iki metotla yapılmaktadır. Bunlar; 1) Bireylerin sayım anında bulunduğu bölge nüfusunda sayılmasıdır.2) Bireylerin sürekli oturma yerlerindeki ikametlerinde (adrese dayalı) nüfusta sayılması metodudur.

Yıl Ortası Nüfus (YON)

Yılın ortası olarak 30 Haziran veya 1 Temmuz olarak kabul edilir. Bir ülkenin veya bölgenin 30 Haziran ya da 1 Temmuz tarihindeki nüfusuna yıl ortası nüfus denir.

Yıl ortası nüfus farklı şekillerde hesaplanmaktadır. En yaygın yıl ortası nüfus hesaplama metodu aşağıdaki gibidir.

$$YON = \text{Yılbaşı Nüfus} + \frac{(\text{Doğumlar} + \text{İçe Göç}) - (\text{Ölümler} + \text{Dışa Göç})}{2}$$

Nüfus Yoğunluğu

Bir bölgede veya ülkede kilometre kare başına düşen birey sayısını gösteren bir ölçüdür. Nüfus yoğunluğunu hesaplamada izdüşüm yüzölçümü kullanılmaktadır.

$$\text{Nüfus Yoğunluğu} = \frac{\text{Nüfus}}{\text{Yüzölçüm}} \text{ biçiminde hesaplanır.}$$

Nüfus Artış Hızı (NAH)

Nüfus artış hızı nüfus istatistikleri içinde en yaygın kullanılan bu ölçü aşağıdaki gibi hesaplanır.

$$NAH = \left(\frac{\text{Yıl içi canlı doğum sayısı} - \text{Yıl içi ölüm sayısı}}{\text{Yıl ortası nüfus}} \right) \times 1000$$

Nüfus ve Nüfus Artış Hızı Tahmin Metotları

Nüfusun belirli bir süre sonra ne kadar olacağını tahmin etmek amacıyla veya nüfusun artışı hızını tahmin etmek için yararlanılan metotlardır. En sık kullanılan nüfus tahmin metotları Üssel Fonksiyon ve Bileşik Faiz ve metotlarıdır.

TUİK (Türkiye İstatistik Kurumu) nüfus tahmini ve nüfus artış hızını tahmin etmede üssel fonksiyon metodunu kullanmaktadır.

Yaşa özel cinsiyet oranı (YÖCO)

Ayrıca ülke veya bölge nüfusunda belirli bir yaş grubu için Yaşa Özel Cinsiyet Oranı (YÖCO) da hesaplanabilir. Bu ölçünün hesaplanmasında;

$$\text{Yaşa Özel Cinsiyet Oranı} = \frac{\text{X yaştaki erkek nüfus}}{\text{X yaştaki kadın nüfus}} \times 100 \text{ formülü kullanılır.}$$

Nüfus Piramidi

Nüfusun, cinsiyete göre yaş gruplarına göre değerlendiren bir grafikdir. Bu grafikte dik koordinat sisteminin y ekseninde beşerli yaş gruplarına göre yaş sınıfları yerleştirilir. Koordinat sisteminde x ekseninin sağ tarafına erkek, sol tarafına kadın nüfusu (sağ taraf kadın sol taraf erkek nüfus içinde düzenlenebilir) benzer ölçekle (sayı veya yüzde) yerleştirilerek çizilen bir grafikdir.

Nüfus piramitleri o ülke veya bölgenin,

Nüfus sayısı,
Cinsiyet durumu,
Nüfusun yaş gruplarına dağılımı,
Doğum ve Ölüm oranları,
Ortalama yaşam süresi,
Nüfustaki hareketlenmeler hakkında bilgi verir.

Genç Bağımlı Nüfus: On beş (15) yaşın altındaki nüfus genç bağımlı nüfus olarak kabul edilmektedir.

Yaşlı Bağımlı Nüfus: Atmış dört (64) yaşın üstündeki nüfus yaşlı bağımlı nüfus olarak tanımlanmaktadır.

Çalışan Nüfus: Çalışma çağındaki 15-64 yaş arası nüfus çalışan nüfustur.

Doğumlar İle İlgili İstatistikler

Bir bölgenin veya ülkenin nüfus yapısındaki değişimi etkileyen en önemli faktör doğumlardır.

Canlı Doğum: Gebelik süresine bakılmaksızın ana vücudundan ayrıldığı anda canlılık belirtisi gösteren (soluk alma, kalp atımı, kordonda nabız ya da çizgili kasların hareketi gibi) gebelik ürününe canlı doğum denir.

Fetüs Ölümü: Gebelik süresine bakılmaksızın ana vücudundan ayrıldığı anda canlılık belirtisi göstermeyen gebelik ürününe fetüs ölümü denir.

Bu bağlamda gebelik süresi, 28 haftadan az olan fetüs ölümüne erken fetüs ölümü ya da düşük, 28 haftadan geç olan fetüs ölümüne geç fetüs ölümü ya da ölü doğum denir.

Doğum ile ilgili istatistikleri yapabilmek ve ölçüler için bazı tanımlamaları bilmek gereklidir. Bu nedenle gebelik süresi ve bebek doğum ağırlığı esas alınarak bazı tanımlamalar yapılmıştır.

Doğum Hızı Ölçüleri

Kaba doğum hızı (KDH):

Bir yıl içinde olan canlı doğum sayısı, yıl ortası nüfusa bölünerek bölüm değeri 1000 ile çarpıldığında "Kaba Doğum Hızı" elde edilir. Bu ölçü doğurganlık hakkında kabaca bilgi verir ve her bin nüfusa kaç canlı doğum düştüğünü gösterir.

$$\text{Kaba Doğum Hızı (KDH)} = \frac{\text{Yıl içindeki canlı doğum sayısı}}{\text{Yıl ortası nüfus}} \times 1000$$

Genel doğurganlık hızı (GDH):

Bu ölçü doğurgan çağda olduğu kabul edilen 15-44 yaş grubundaki (Yaş grubu 15-49 olarak ta alınabilmektedir) her bin kadına düşen canlı doğum sayısını gösterir ve aşağıdaki gibi hesaplanır.

$$\text{GDH} = \frac{\text{Yıl içi canlı doğum sayısı}}{\text{15 - 44 yaş grubu kadın yıl ortası nüfusu}} \times 1000$$

Toplam doğurganlık hızı

15-49 yaş grubundaki kadınlar için ortalama çocuk sayısını veren bir ölçüdür. Toplam doğurganlık hızı 2010 yılı için Türkiye geneli olarak 2.03 tür.

Ölümler İle İlgili İstatistikler

Toplumun nüfus yapısındaki değişimi etkileyen bir diğer faktör ölümlerdir. Ölüm olgularının en önemlil ölçüleri; kaba ölüm hızı, bebek ölüm hızı, yaşa veya cinsiyete özel ölüm hızı ölçüleridir.

Kaba ölüm hızı (KÖH)

Bir toplum veya bölgede ölüm olguları hakkında genel bir bilgi veren bir göstergedir. İncelenen toplumda bir yıl içinde görülen toplam ölüm sayısının yıl ortası nüfusa bölünmesiyle elde edilen bir hızdır. Bölüm değeri 1000 ile çarpılarak binde olarak ifade edilir. Ayrıca yaş grubuna özel ölüm oranı (YGÖO) da hesaplanmaktadır.

Mortalite (Bir hastalıktan ölüm) hızı:

İncelenen toplumda (bir il, bölge, ülke) belirli bir hastalıktan ölümlerin o toplumun yıl ortası nüfusuna bölünmesi ile bulunan bir ölüm hızıdır. Bölüm değeri 100000 ile çarpılarak bu ölçü genelde yüzbinde olarak hesaplanır.

$$\text{Mortalite Hızı} = \frac{\text{X Hastalığın dan ölen birey sayısı}}{\text{Yıl ortası Nüfus}} \times 100000$$

Bebek ölüm hızı (BÖH):

BÖH, bebek ölüm oranı olarak ta ifade edilmektedir. BÖH, bir toplumun, ülkenin ya da bölgenin sağlık ve gelişmişlik düzeyini gösteren önemli göstergelerden biri olarak kabul edilmektedir. Bu

ölçü aşağıdaki gibi hesaplanır.

$$\text{BOH} = \frac{\text{Yıl içinde bir yaşını doldurmada n ölen bebek sayısı}}{\text{Bil yılda doğan canlı doğum sayısı}} \times 1000$$

DERS ADI: Biyoistatistik

ÜNİTE ADI: Sağlık Alanına Özel İstatistikler

ÜNİTE NO: 14

YAZAR: Prof. Dr. ÖMER AKBULUT

Sağlık alanında özel istatistikler genellikle hastalıklarla ilgili istatistiklerdir. Hastalıklarla ilgili hız ölçüleri, medikal tanı testlerinin güvenilirlik ölçüleri ve hastalıklar ile hastalık nedenleri arasındaki bağımlılığı ifade eden oransal ölçüler sağlık alanına özel istatistiklerin en önemlileridir.

Hastalık Hızı İle İlgili İstatistikler

Hastalık hızları, prevalans ve insidans hızı olarak iki şekilde incelenir.

Prevalans Hızı

İnceleme süresi içinde eski ve yeni olguların tümü dahil mevcut hasta sayısının risk altındaki nüfusa bölünmesiyle elde edilen bir hastalık hızı (oranı) ölçüsüdür.

$$\text{Prevalans Hızı} = \frac{\text{İnceleme süresi içinde saptanan olgu sayısı}}{\text{Risk altındaki nüfus}} \times k$$

Burada (k) olgu sayısına göre 100, 1000, veya ... 100000 olarak alınır.

Prevalans hızı, genelde kronik ve kanser türü hastalıkların görülme oranlarını ifade etmekte kullanılan hızdır.

İnsidans Hızı

İncelenen sürede yeni gözlenen olgu sayısının risk altındaki nüfusa bölünmesiyle elde edilen yeni olgu gözlenme hızıdır. Aşağıdaki gibi hesaplanır.

$$\text{İnsidans Hızı} = \frac{\text{İnceleme süresi içinde saptanan yeni olgu sayısı}}{\text{Risk altındaki nüfus}} \times k$$

Burada yeni olgu sayısı dikkate alınarak k; 100, 1000 veya 100000 olarak alınır. İnsidans hızı hesaplanmasında inceleme süresi içindeki yeni olgu sayısı dikkate alınırken Prevalans hızı hesaplanmasında hem yeni hem eski mevcut tüm olguların işleme alındığına dikkat edilmelidir. İnsidans hızı daha çok akut ve bulaşıcı hastalıkların değerlendirilmesinde kullanılmaktadır.

Fatalite Hızı (Bir Hastalığın Öldürücülük Hızı)

Belirli bir süre içinde A hastalığına yakalananların bu süre içinde ne kadarının öldüğünü ifade etmek için kullanılır. Hastalığın türüne göre belirlenir ve aşağıdaki gibi hesaplanır.

$$\text{Fatalite Hızı} = \frac{\text{A hastalığından ölüm sayısı}}{\text{A hastalığına yakalanan kişi sayısı}} \times k$$

Mortalite hızı ve Fatalite hızlarının ayırımının iyi yapılması gerekir. Her ikisinde ölüm hızı ile ilgilidir. Mortalite hızı toplumdaki ölüm hızını gösteren bir ölçüdür. Fatalite hızı bir hastalığın öldürücülük hızıdır. İncelenen hastalığa özel bir ölçüdür.

Medikal Tanı Testlerinin Güvenirlik Ölçüleri

Sağlık bilimlerinde medikal testlerin güvenilirliğinin özel bir önemi vardır. Medikal tanı testi veya aracı, bir referansa göre değerlendirilir. Referans geçerliği ve güvenilirliği kanıtlanmış bir metottur. Bir test, gerçekten hastalık var ise hastalığı tanımalı, hastalık yok ise hastalığın olmadığını belirlemelidir. Genelde patolojik inceleme sonucu veya klinikteki gözlem kesin sonuç (referans) olarak kabul edilir. Bir medikal testin veya aracın gerçek durumu yansıtması aşağıdaki tablodaki gibi özetlenmektedir.

Tablo Medikal Testin Gerçek Durumu Yansıtması

		Gerçek Durum (Referans)		
		H +	H-	Toplam
Medikal Test Sonucu	H+	Gerçek Pozitif (GP=a)	Yalancı Pozitif (YP=b)	T(T+)
	H-	Yalancı Negatif (YN=c)	Gerçek Negatif (GN=d)	T(T-)
Toplam		G(H+)=a+c	G(H-)=b+d	N

Medikal testlerin değerlendirilmesinde aşağıdaki ölçüler kullanılır.

Duyarlılık, D (Sensitivity) Oranı

Gerçekten hasta olan (G(H+)) bireylerin test tarafından hangi oranda tespit edildiğini belirten bir ihtimal (olasılık) değeridir. Aşağıdaki gibi hesaplanır.

$$\text{Duyarlılık (Sensitivity)}=P(D)= a / (a+c) \text{ veya Duyarlılık} = GP / (GP+YN)$$

Özgüllük, Ö (Specifity) Oranı

Bir testin gerçekten hasta olmayanları ayırabilme yeteneğini belirten orandır. Özgüllük, testin ölçmek istediği A olayını başka olaylarla karıştırmaksızın tanıyıp tanımadığını belirtir.

$$\text{Özgüllük (Specifity)}= P(\bar{O})= d / (b+d) \text{ veya Özgüllük} = GN / (YP+GN)$$

Pozitif Tanımlama Oranı, PTO (Positive Predictivity)

Bir testin gerçekten hasta diye nitelediği kişilerin gerçekten ne kadarının hasta olduklarını gösteren bir orandır. Bu ölçü aşağıdaki gibi hesaplanır.

$$\text{Pozitif Tanımlama Oranı}=P(PP)=a / (a+b) \text{ veya PTO} = GP / (GP+YP)$$

Negatif Tanımlama Oranı, NTO (Negative Predictivity)

Bu ölçü negatif kestirim oranı olarak ta adlandırılmaktadır. Testin hasta olmadığını belirttiği kişilerin gerçekte hangi oranda hasta olmadıklarını gösteren bir orandır ve aşağıdaki gibi hesaplanır.

$$\text{Negatif Tanımlama Oranı}=P(NP)=d / (c+d) \text{ veya NTO} = GN / (YN+GN)$$

Doğruluk Oranı, DO (Diagnostic Value)

Bu ölçünün bir diğer adı "Tanı değeri"dir. A hastalığına ilişkin bir medikal testin, bir bireyi gerçekten hasta ise hasta (H+), değilse (H-) hasta olmadığını tanımlama oranıdır. Aşağıdaki gibi hesaplanır.

$$\text{Doğruluk Oranı}=P(DO)=(a+d) / (N) \text{ veya DO} = (GP+GN) / N$$

Bir testin Özgüllük, Duyarlılık ve Doğruluk Oranı (Tanı değeri) büyük önem taşır. Eğer yanlış hastalık teşhisi daha ileri tekniklerle düzeltilemiyorsa testin özgüllük oranının yüksek olması istenir. Eğer test, tarama amacıyla kullanılıyorsa Duyarlılık oranının yüksek olması istenir. Tanı değeri yüksek testler hem klinikte tanı amacıyla hem de taramalarda güvenle kullanılabilir.

Hastalık Nedenlerine İlişkin Oranlar

Hastalık ile hastalık nedenleri arasındaki bağımlılığı ortaya koymak için geliştirilmiş oranlar mevcuttur. Bu ilişkinin ortaya konabilmesi için hastalık ve hastalık nedenlerinin nominal (adlandırma) olarak belirlenmesi gerekir. Bu ilişkiyi ortaya koymak için Risk Oranı ve Odds Oranı olmak üzere iki oran geliştirilmiştir.

Hastalık varlığı veya yokluğu ile hastalık nedeninin varlığı ve yokluğu 2x2 boyutlu bir tabloda aşağıdaki gibi özetlenebilir

Tablo Hastalık ve Hastalık Nedeni İlişkisi Tablosu

		Hastalık Durumu A		Toplam
		Hasta (A+)	Hasta Değil (A-)	
Hastalık nedeni (X)	Var (X +)	a= (X+ve A+)	b= (X+ve A-)	R1= X+
	Yok (X -)	c= (X-ve A+)	d= (X-ve A-)	R2= X-
Toplam		C1=(A+)	C2=(A-)	N

Risk Oranı

Relatif Risk olarak da adlandırılan bu oran, etki altında kalan birey insidansının etki altında kalmayan birey insidansına oranı ile hesaplanan bir ölçüdür. Daha çok "İleriye yönelik çalışma düzenlerinde" veriler yukarıdaki şartlara uyuyor ise Risk Oranı (RR, Relatif Risk) kullanılır. Risk Oranı tablodaki semboller kullanılarak aşağıdaki gibi hesaplanır.

$$RR= \frac{a / (a + b)}{c / (c + d)} \text{ şeklinde hesaplanır.}$$

Odds Oranı

Odds oranı (OR, Odds Ratio) daha çok retrospektif, çalışmalarda kullanılır. Odds oranı, her grupta hastalık nedeni altında olanların hastalık nedeni altında olmayanlara oranıdır. Odds oranı ise,

$$OR= \frac{a / c}{b / d} = \frac{a * d}{b * c} \text{ şeklinde hesaplanır.}$$

