**:iSTATiSTiK:**

**Ders01 - İstatistik İle İlgili Bazı Temel Kavramlar**

**İstatistik Nedir?:**

İstatistik, en genel yaklaşımla, 'sayısal bilgi', 'sayısal veri' ya da 'ham veri' olarak tanımlanmaktadır. Örneğin 'Ayşe, 45kg ağırlığındadır' ifadesindeki 45kg, Ayşe'nin 'ağırlık' olarak tanımlanan bireysel hakkında bir bilgi vermektedir ve bir istatistiktir.

Bir diğer yaklaşıma göre istatistik, bir süreç tanımlamaktadır. Bir süreç olarak istatistik, bir veri seti üzerinde yapılan aşamalı işlemleri ifade etmektedir. Veriler üzerinde, temelde betimleme ve çıkarsama işlemleri yapılabilmektedir.

**Gözlem:**

İstatistik özelinde gözlem; belli bir özellik hakkında bilgi ya da veri toplamak için yürütülen amaçlı ve iyi tanımlı bir eylemdir. Özellikler, 'doğrudan gözlenebilir özellikler' ve 'dolaylı gözlenebilir özellikler' olmak üzere iki grupta değerlendirilebilmektedir.

**Evren - Örneklem:**

Evren (population), gözlenmesi amaçlanan belli bir özelliğe sahip gözlem birimlerinin tamamını içeren en geniş gözlem kümesi olarak tanımlanır. Evren, 'ana kitle', 'ana kütle' ya da 'popülasyon' olarak da ifade edilebilmektedir.

Örneklem (sample), belli bir özelliğin üzerinde gözlendiği ve evrenin bir alt kümesi olan gözlem birimleri kümesidir. Tanımlanmış bir evrenden bir örneklem belirleme işlemelerini içeren yöntem ve tekniklere 'örnekleme (sampling)' denir. Tanımdan da anlaşılacağı üzere, örneklem büyüklüğü yani örnekleme dahil olan gözlem birimleri sayısı, evrenden küçüktür. Ayrıca bir evrenin içerisinde birden çok sayıda alt küme bulunabileceği için birden çok sayıda da örneklem belirlenebilir.

**Homojenlik - Heterojenlik:**

Gözlenen bir özellik açısından birbirine yakın ya da benzer gözlem birimlerinden oluşan gruplara 'homojen grup' denir. Aksi durumda gözlenen özellik açısından birbirinden farklı gözlem birimlerinden oluşan gruplara 'heterojen grup' denir. 'Tam homojen grup' denildiğinde, her bir grup üyelerinin her birinin aynı özellikte olduğunu, 'tam heterojen grup' ise grup üyelerinin her birinin farklı özellikte olduğunu ifade eder.

Örneğin, "akademik başarı açısından A şubesi B şubesinden daha homojendir" denildiğinde, ilgilenilen özelliğin akademik başarı olduğu, A şubesindeki öğrencilerin akademik başarı ortalamalarının ya da notlarının B şubesine göre birbirine daha yakın olduğu anlaşılır.

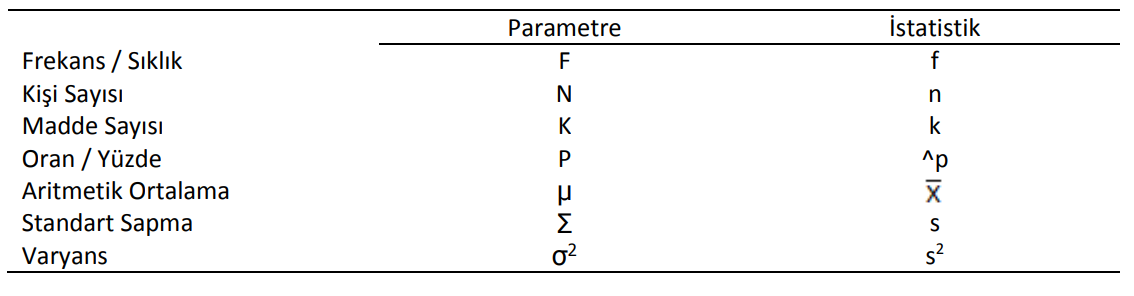
**Parametre - İstatistik:**

Alana özel kullanımıyla parametre, evrenden doğrudan elde edilen sayısal değerler ya da kestirimlerdir. 'Evren değer' olarak da ifade edilir. İstatistik ise örneklemden elde edilen sayısal

değerler ya da kestirimlerdir. 'Örneklem değer' olarak da ifade edilir.

Bu tanımlamalara göre istatistiklerle parametrelerin olasılıklı olarak örtüşme düzeylerinin test edilmesinde kullanılan yöntem ve teknikler de istatistiksel yöntem ve teknikler olarak tanımlanabilmektedir.

Evren değer yani parametreler ile örneklem değer yani istatistikler, farklı sembollerle gösterilebilmektedir.



**Değişken Türleri:**

1. **Nitel Değişken - Nicel Değişken:**

Bazı değişkenler, sembollerle ya da sözcüklerle ifade edilir. Bazı değişkenler ise sayılarla ifade edilir. Sözcük ya da sembollerle ifade edilen değişkenlere 'nitel (qualitative)', sayılarla ifade edilen değişkenler ise 'nicel (quantitative)' değişken olarak adlandırılır.

Örneğin cinsiyet, kadın ve erkek olarak değerler alabilen iki kategorili ve nitel bir değişkendir. Bir başka örnek olarak eğitim düzeyi, okuryazar değil, okuryazar, ilkokul mezunu, ortaokul mezunu gibi değerler alabilen çok kategorili ve nitel bir değişkendir.

Yaş, ağırlık, öğrenme başarısı, akademik aşarı, sıcaklık, aylık gelir gibi değişkenler ise sayılarla ifade edilir ve bunlar nicel değişkenlerdir.

2. **Sürekli Değişken - Süreksiz/Kesikli Değişken:**

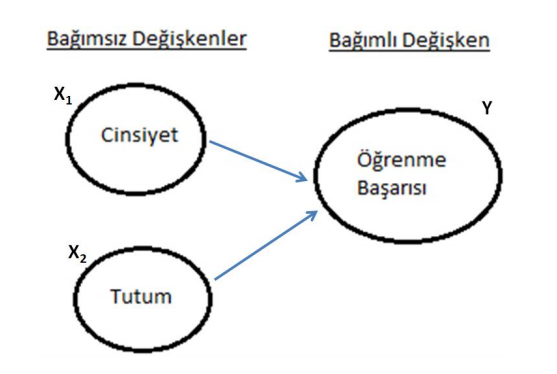
Matematiksel bir kavram olarak süreklilik, herhangi iki sayının ortasında her zaman bir üçüncü sayının bulunması durumunu ifade etmektedir. Örneğin doğal sayılar ve tamsayılar kümeleri, sürekli sayılar değildir. Rasyonel sayılar kümesi ise süreklidir.

Cinsiyet, eğitim düzeyi, doğum yeri, kan grubu, şube gibi değişkenler kesikli, yaş, ağırlık, akademik başarı, öğrenme başarısı, tutum gibi değişkenler süreklidir. Genellikle nitel değişkenler kesikli, nicel değişkenler ise sürekli değişkenlerdir.

3. **Bağımlı Değişken - Bağımsız Değişken:**

Herhangi bir değişken, tek başına bağımlı ya da bağımsız olarak sınıflandırılamaz. Bu sınıflamanın yapılabilmesi için en az iki değişken ve bu değişkenler arasında tanımlı bir ilişki bulunmalıdır.

İki değişkenden biri diğer değişkenle ilişkilendirildiğinde, ilişkinin yönüne göre, bağımlı-bağımsız sınıflaması yapılabilir. İlişkili iki değişkenden biri diğeri aracılığı ile (yani diğerine bağlı olarak) betimlenecek ya da açıklanacak ise açıklanan temel değişken 'bağımlı değişken (dependent variable)', bu değişkenle ilişkilendirilen diğer değişken ise 'bağımsız değişken (independent variable)' olarak adlandırılır.



**Ölçek ve Ölçekleme:**

Özel kullanımıyla ölçek, genellikle tutum, ilgi, kişilik gibi duyuşsal özelliklerin gözlenmesinde ya da ölçülmesinde kullanılan özel bir veri toplama aracıdır. Diğer bir tanımlamayla ölçek, belli birimlerle birimlendirilmiş ve yapılandırılmış bir ölçme aracıdır. Bu özel kullanımının yanı sıra bir değişkenin alabileceği olası değerlere de ölçek denilmektedir. Bir değişkenin sayı ya da sembol olarak alabileceği değerler, bir ölçek üzerinde ifade edilmektedir.

Bu değerlerin yani bir değişkenin alabileceği değerlerin belirlenmesi, değişkenin seçeneklerinin belirlenmesi işlemlerine de 'ölçekleme' denilmektedir. Bir değişkenin farklı şekillerde ölçeklenmesi mümkündür. Ölçekleme, toplanacak verinin niteliğini de etkilemektedir.

Örneğin bir değişken olarak yaş değişkenini dikkate alalım. Bireylerin yaşları hakkındaki bilgileri farklı şekillerde almak mümkündür. Örneğin "Yaşınız:...." şeklinde sorulan açık uçlu bir soruya, 24, 25, 34, 38 gibi verilen yanıtlarla yaş değişkenine yönelik veriler sürekli ve nicel olarak toplanabilir. Aynı değişkene yönelik veriler "Yaşınız: ( ) 20 ve altında, ( ) 21-25 arası, ( ) 26-30 arası, ( ) 31 ve üzeri" şeklinde sorulan bir soru ile de toplanabilir. Bu durumda elde edilen veriler, ilk duruma göre daha sınırlı olacaktır.

**Ölçek Türleri:**

1. Sınıflama Ölçeği (Nominal scale) 2. Sıralama Ölçeği (Ordinal scale)

3. Eşit Aralıklı Ölçek (Interval scale) 4. Eşit Oranlı Ölçek (Ratio scale)

Sınıflama ölçeği düzeyindeki bir değişken, sadece grup ya da gruba aidiyet hakkında bilgi verir. Örneğin kan grubu, doğum yeri, cinsiyet, şube, medeni durum gibi değişkenler sınıflama ölçeği düzeyinde değişkenlerdir.

Sıralama ölçeği düzeyindeki değişkenler, gruba aidiyet bilgisi yanı sıra kategoriler arasında bir sıra ilişkisi de gösterir. Örneğin anne eğitim düzeyi, sıralama ölçeğinde bir değişkendir. Örneğin lise mezunu olman bir anne, ortaokul mezunu olan bir anneden daha yüksek bir eğitim düzeyine sahiptir.

Eşit aralıklı ölçek düzeyindeki bir değişkende değişkenin alabileceği ardışık değerler arasında eşit mesafe vardır. Yani bu değişkenin değerlerinin yer aldığı ölçek, eşit aralıklıdır. Örneğin öğrenme başarısının bir göstergesi olarak 'not' şeklinde tanımlanan değişken, 0 ile 100 arasındaki notlardan oluşuyorsa bu değişken eşit aralık ölçeğindedir. Yani 1 notu ile 2 notu, 2 notu ile 3 notu ve benzeri arasındaki mesafeler birbirine eşittir.

Eşit oranlı ölçekte düzeyindeki bir değişkende de ölçekleme, eşit aralıklıdır. Fakat eşit aralık ölçeğinden farkı, 'gerçek/mutlak sıfır' noktasına sahip olmasıdır. Örneğin ağırlık, eşit oranlı ölçek düzeyinde bir değişkendir. 0kg, ağırlığın olmadığını yani mutlak yokluğu ifade eder.

Ölçek türleri, bir değişkenin değerleri üzerinde yapılabilecek işlemlerin belirlenmesinde yardımcı olur.

* Sınıflama (Nominal) ölçeği düzeyindeki değişkenlerde basit sayma ve sınıflama işlemleri yapılabilir.
* Sıralama (Ordinal) ölçeği düzeyindeki değişkenlerde sayma ve sınıflama işlemlerinin yanı sıra medyan, sıra farkları ve korelasyon hesaplamaları yapılabilir.
* Eşit aralıklı (Interval) ölçek düzeyindeki değişkenlerde, toplam ve farka dayalı istatistiksel hesaplamaların tamamı kullanılabilir.
* Eşit oranlı (Ratio) ölçek düzeyinde ise çarpma ve bölme işlemleri dahil tüm işlemleri içeren istatistiksel hesaplamalar kullanılabilir.

**Ders02 - Veri Setinin Hazırlanması**

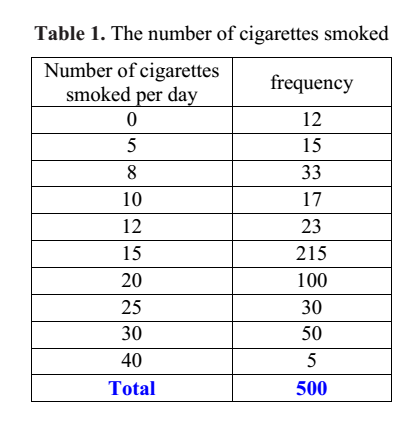
**Frekans Tabloları:**

Frekans, kısaca görülme sıklığıdır. Bir gözlem değerinin ya da ölçme sonucunun tüm gözlem birimleri içerisinde görülme sıklığı, bu değerin frekansı olarak belirlenir. Örneğin 30 kişilik bir sınıfta öğrencilerin öğrenme başarısını ölçmek için bir yazılı yoklama yapıldığını ve 0 ile 100 arasında notlar verildiğini düşünelim. 6 öğrenci 70, 8 öğrenci 65, diğer öğrenciler ise başkaca notlar almış olsun. Bu durumda 70 değerinin görülme sıklığı yani frekansı 6, 65 notunun görülme sıklığı yani frekansı 65 olmuş olur.

Frekans tabloları, bir ya da bir kaç değişkene yönelik frekansların, yapılandırılmış bir şekilde sunumudur.

Frekans tabloları tek bir değişkene yönelik olarak hazırlanabileceği gibi iki veya daha fazla değişken için de hazırlanabilir. Bir frekans tablosunda betimlenen değişken sayısı arttıkça, tabloda yer alan bilgilerin anlaşılması ve yorumlanması da o ölçüde güçleşir.

Frekans Tablosu:



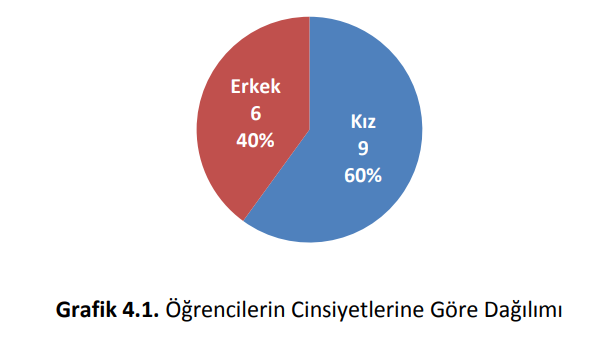
**Frekans Grafikleri:**

Grafiklerin, frekans dağılımlarının betimlenmesi amacıyla kullanılmasında dikkat edilmesi gereken önemli noktalardan biri, değişkenin türüne ve karakteristiğine göre kullanılacak uygun grafik türünün seçilmesidir. Bu bölümde frekans dağılımlarının betimlenmesinde yaygın olarak kullanılan 4 grafik türü tanıtılmaktadır.

1. **Pasta Grafiği (Pie Chart):**

'Daire grafiği' olarak da bilinen pasta grafiği, 360 derecelik bir dairenin dilimleri ile frekans dağılımlarının gösterildiği bir grafik türüdür. Dilimlerin büyüklüğü frekans yüzdeleri ile orantılı olacak şekilde belirlenir. Pasta grafikleri, cinsiyet, şube, kan grubu gibi kesikli ve kategorik değişkenlerin frekans dağılımlarının gösterilmesinde kullanışlıdır. Değişkenlerin kategori sayısı arttıkça, pasta grafiği daha karışık bir şekle dönüşecektir. Bu nedenle pasta grafiği, 2, 3, 4 gibi az kategorili değişkenlerde, daha fazla çok kategorisi olan değişkenlere göre daha kullanışlıdır.

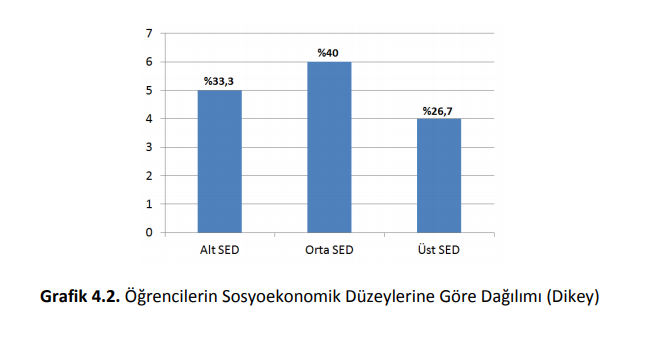
CINS: K; K; K; E; E; K; E; E; K; K; E; K; K; E; K



1. **Sütun Grafiği (Bar Chart):**

‘Bar grafiği' olarak da bilinen sütun grafiği, frekansların kartezyen sistem üzerinde sütun yükseklikleri ile eşleştirildiği bir gösterim biçimidir. Sütun grafiği 'yatay' ve 'dikey' olmak üzere iki biçimde kullanılabilmektedir. Alışıldık olan gösterim biçimi dikey sütun grafiğidir. Bu grafikte, değişken ve değişkenin kategorileri yatay eksen ile, frekanslar ise dikey eksen ile temsil edilir. Pasta grafiğinde olduğu gibi sütun grafiği de kesikli ve kategorik değişkenlerin frekans dağılımlarının gösterilmesinde kullanışlı bir grafik türüdür. Sütun grafiğinin ayırıcı önemli bir özelliği, sütunlar arasında boşluk bırakılmasıdır. Bu durum, kesikli değişkenlerin yani sınıflama (Nominal) ve sıralama (Ordinal) ölçekleri düzeyindeki veriler için kullanılan bir grafik türü olmasından kaynaklanmaktadır.

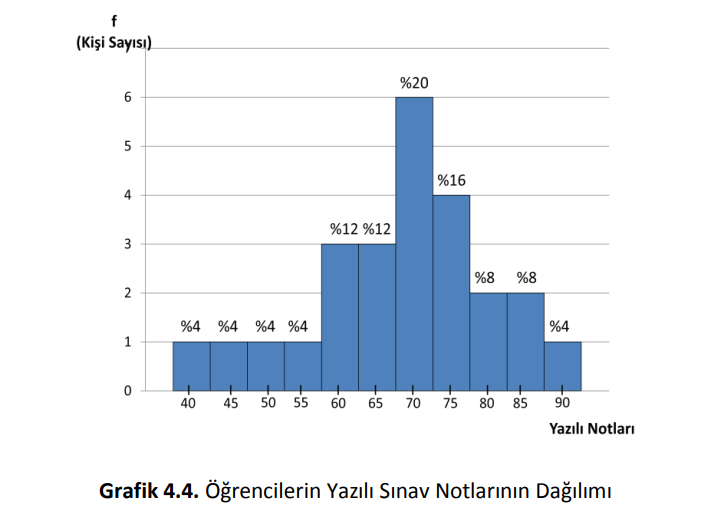
SED (SosyoEkonomik Düzey): Alt; Alt; Orta, Üst; Üst; Orta, Orta; Orta; Alt; Üst; Alt; Alt; Orta; Orta; Üst



**3) Histogram:**

Histogram, sürekli değişkenlerin frekans dağılımlarının betimlenmesinde kullanılan, özel bir sütun grafiğidir. Kesikli verilerde kullanılması daha uygun olan sütun grafiği ile histogram arasındaki temel fark, histogramda sütunlar arasında boşluk bırakılmamasıdır. Bu durum histogramın sürekli değişkenler yani en az eşit aralık ölçeğindeki değişkenlerin frekans dağılımlarının gösterilmesinde kullanılmasıdır. Sütun grafiğine benzer şekilde histogram da kartezyen sistemde çizilmektedir. Dikey eksen frekans ekseni, yatay eksen ise değişken eksenini tanımlamaktadır. Sürekli değişkenlere yönelik frekans dağılımlarında olduğu gibi histogramlarda da gerekli durumlarda ölçme sonuçlarının gruplanarak verilmesi mümkündür. Fakat gruplama, veri kaybına yol açacağı için rasyonel ve kabul edilebilir bir gerekçe olmaksızın gruplama yapılmaması önerilir. Verilerin gruplanarak sunulmasının kabul edilebilir bir gerekçesi, değerlerin çok geniş bir aralıkta yayılma gösteriyor olmasıdır.

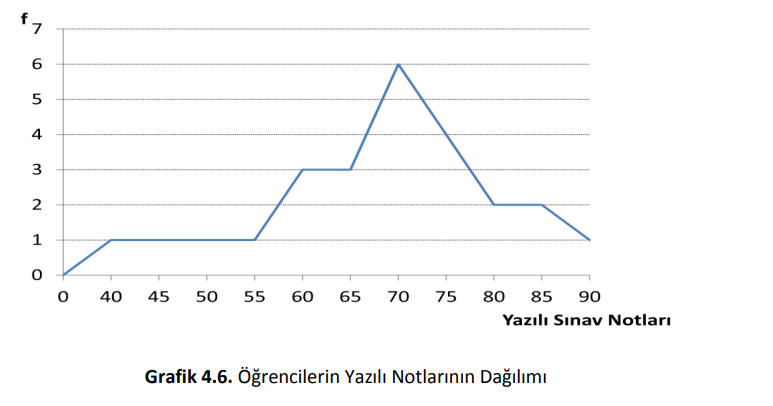
NOT: 55; 60; 65; 75; 90; 95; 90; 80; 75; 75; 70; 65; 60; 50; 45; 40; 70; 65; 70; 70; 60; 70; 80; 75; 70



**4) Çizgi Grafiği (Polygon Chart):**

Çizgi grafiği histograma göre daha kolay çizilebilen ve daha alışıldık bir grafik türüdür. Çizgi grafiği de kartezyen düzlemde çizilir. Yatay eksen değişken ve değişkenin değerlerinin yer aldığı, dikey eksen ise frekans değerlerinin yer aldığı eksenlerdir. Her iki eksendeki değerler de eşit aralıklı olarak eksene yerleştirilir. Çizgi grafiği, yatay eksende ölçme sonuçları yani değişkenin değerleri ile dikey eksende bu değerlerin frekanslarının kesişim noktalarının belirlenmesi ve bu noktaların ardışık olarak düz çizgilerle birleştirilmesi ile elde edilir.

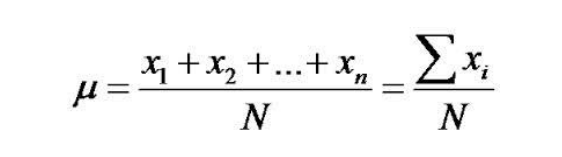
Çizgi grafiği, histogram da olduğu gibi sürekli değişkenlerin frekans dağılımlarının betimlenmesinde kullanılan bir grafik türüdür. Çizgi grafiği süreklilik gösterir. Çizgi grafiği sürekli değişkenlere yönelik olarak sıklıkla kullanılan bir grafik türü olmakla birlikte özellikle gözlem sayısının az olduğu durumlarda yanıltıcı olabilmektedir. Bu nedenle çizgi grafiği kullanmada ve yorumlamada dikkatli olunması gerekmektedir.



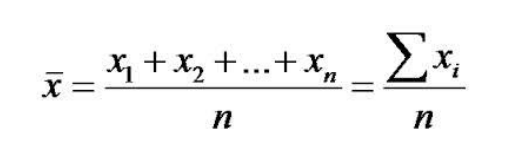
**Ders03 - Merkezi Eğilim Ölçüleri**

**Aritmetik Ortalama (Mean):**

Evrenden elde edilen veriler üzerinde hesaplama yapılıyorsa, ortalama 'evren ortalaması (population mean)' olarak isimlendirilir. Evren ortalamasının formülü aşağıda verilmiştir:



Örneklemden elde edilen veriler üzerinde hesaplama yapılıyorsa ortalama, 'örneklem ortalaması (sample mean)' olarak isimlendirilir. Örneklem ortalamasının formülü aşağıda verilmiştir:



Ortalama, normal dağılım gösteren ve en az eşit aralıklı ölçek düzeyinde olan sürekli verilerde, en 'sağlam' merkezi eğilim ölçüsü olarak bilinir. Bu varsayımları sağlayan verilerde ortalama, verilerin tamamını temsil eden bir sayısal değer olarak dikkate alınabilir.

**Ortanca (Median):**

Medyan (median); küçükten büyüğe doğru sıralanmış verilerin tam ortasında kalan değerdir.

**Tepe Noktası (Mod):**

En çok tekrar eden sayı.

**Yüzdelik ve Çeyrek Sapma:**

Yüzdelik; küçükten büyüğe doğru sıralanmış verilerin belli bir yüzdesini altında bırakan noktadaki gözlenen değerdir. 'Yyüzde' sembolü ile gösterilir. Örneğin;

Y20; sıralanmış verilerin %20'sini aşağısında bırakan sıradaki gözlenen değeri,  Y25; sıralanmış verilerin %25'ini aşağısında bırakan sıradaki gözlenen değeri,

Y50; sıralanmış verilerin %50'sini aşağısında bırakan sıradaki gözlenen değeri,  Y75; sıralanmış verilerin %75'ini aşağısında bırakan sıradaki gözlenen değeri,

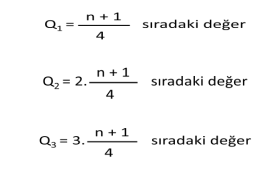
ifade eder.

Çeyreklik (quartile) ise sıralanmış verilerin çeyrek yani %25'lik dilimlerinin başlangıç ve bitiş noktalarında yer alan değerleri ifade eder. Dolayısıyla çeyreklik değerler aynı zamanda yüzdelik değerlerle ifade edilebilir. Çeyreklik, 'Q' sembolü ile ifade edilir. Teorik olarak, sıralanmış bir veri setinde 4 çeyrek bölüm vardır. Dördüncü çeyreklik, verilerin %100'ünün yani tamamının bir üstündeki değeri ifade eder. Bu nedenle pratikte dördüncü çeyreklik kullanılmaz.

Birinci çeyreklik Q1, ikinci çeyreklik Q2 üçüncü çeyreklik Q3 sembolleri ile gösterilir. Birinci çeyreklik 25. yüzdelik ile ikinci çeyreklik 50. yüzdelik ile üçüncü çeyreklik 75. yüzdelik ile örtüşür.

Q1=Y25 Q2=Y50 Q3=Y75

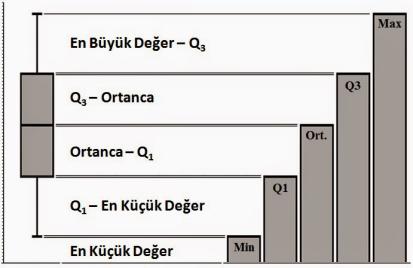
Çeyreklik değerlerin hesaplanmasında hangi sıradaki gözlenen değerin çeyreklik olduğunun belirlenmesinde aşağıdaki formüller kullanılabilir:



Yüzdelik ve çeyreklik, en az eşit aralıklı ölçek düzeyindeki değişkenlerde anlamlı bir istatistiktir. Cinsiyet, medeni durum, şube gibi sınıflama ölçeği düzeyindeki değişkenlerde bu istatistik anlamlı değildir. Eğitim düzeyi, sınıf gibi sıralama ölçeği düzeyindeki değişkenlerde ise yüzdelik ve çeyreklik yerine 'sıra farkları (rank)' kullanılmaktadır.

**Kutu Grafiği (Box Plot):**

İstatistik biliminde kutu grafiği bir betimsel istatistik ve istatistiksel grafik aleti olup niceliksel verileri görsel şekilde özetlemek için Amerikan istatistikçi John Tukey tarafından kutu-ve-bıyıklar grafiği adı altında bir açıklayıcı veri analizi aracı olarak ilk defa geliştirilmiştir.

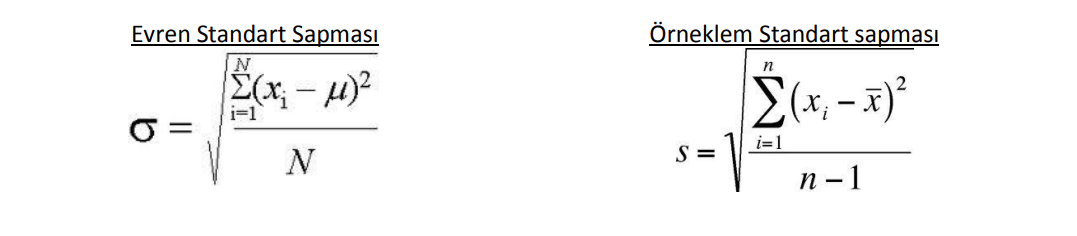


**Ders04 - Merkezden Dağılım Ölçüleri**

**Standart Sapma:**

'Standart kayma' olarak da ifade edilen standart sapma (standard deviation); bir veri setinde her bir verinin ortalamadan uzaklıklarının standartlaştırılmış bir ölçüsüdür.

Evren standart sapması ve örneklem standart sapması, farklı sembollerle ifade edilmenin yanı sıra farklı formüllerle hesaplanır.



Formüllerde yer alan;

 Xi, her bir gözlem birimine yönelik ölçme sonucunu

 µ, evren ortalamasını,

 , örneklem ortalamasını,

 n, örneklem büyüklüğünü yani örneklemde yer alan nesne yada kişi sayısını,

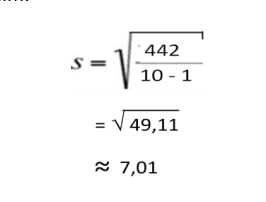
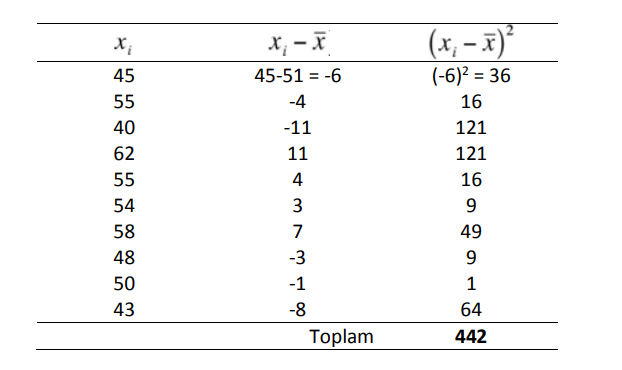
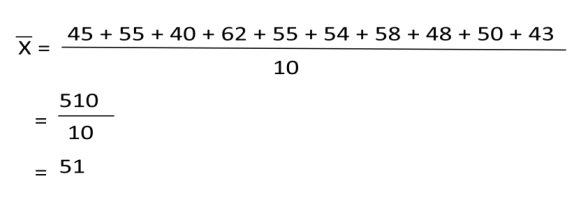
 N, evren büyüklüğünü yani evrende yer alan gözlem birimi sayısını ifade etmektedir.

Dikkat edilirse örneklem standart sapmasının belirlenmesinde (n-1)'e bölme işlemi yapılmaktadır. Bu değer 'serbestlik derecesi' olarak isimlendirilir. Daha az yanlı yani daha titiz bir kestirim sağlamaktadır.

Standart sapma, ortalamaya bağlı olarak hesaplanan bir istatistiktir. Ortalamada olduğu gibi standart sapma da sürekli değişkenlerde ve özellikle normal dağılım gösteren değişkenlerde kullanışlıdır. Değişkenin sürekli yani en az eşit aralık ölçeğinde olduğu ve değişkenin gözlenen değerlerinin normal dağılım gösterdiği durumlarda standart sapma en 'sağlam' merkezden dağılma ölçüsü olarak bilinir. Cinsiyet, eğitim düzeyi, doğum yeri, sınıf, şube, okul türü gibi kategorik değişkenlerde yani sınıflama (nominal) ve sıralama (ordinal) ölçeği düzeyindeki değişkenlerde, ortalama anlamlı olmadığı gibi standart sapma da anlamlı değildir.

Verilerin normal dağılımdan sapması ya da uç noktaların etkisinin söz konusu olduğu durumlarda, ortalamada olduğu gibi standart sapma da doğrudan yorumlanabilir değildir. Bu durumlarda standart sapma, yanıltıcı olabilir, eksik ve yanlış betimsel yorumlara yol açabilir. Örneğin;

Ağırlık: 45kg; 55kg; 40kg; 62kg; 55kg; 54 kg; 58kg; 48kg; 50kg; 43kg



**Varyans:**

Varyans, standart sapmanın karesidir. Standart sapma gibi varyans da verilerin ortalamadan uzaklaşma düzeyleri yani verilerin çeşitliliği hakkında bilgi verir.

Evren varyansı ve örneklem varyansı, farklı sembollerle gösterilmenin yanı sıra hesaplama formüller açısından da farklıdır:



Varyans da sürekli ve normal dağılım gösteren değişkenlerde 'sağlam' bir kestirimdir. Bu varsayımları karşılamayan değişkenlerde varyans, ya anlamsızdır ya da yanıltıcı olabilir.

**Ranj (Range):**

Ranj (range); ölçme sonuçlarını gösteren bir veri setindeki alt ve üst değerlerin yani maksimum ve minimum değerlerin farkıdır. Yani bir değişkene yönelik gözlenen değerlerin aralığıdır.

Ranj = Maksimum Gözlenen Değer - Minimum Gözlenen Değer

**Çeyrek Sapma:**

Çeyrek sapma (quartile deviation); birinci çeyreklik ile üçüncü çeyreklik değerlerinin farkının yarısı alınarak hesaplanan bir merkezi dağılım ölçüsüdür.

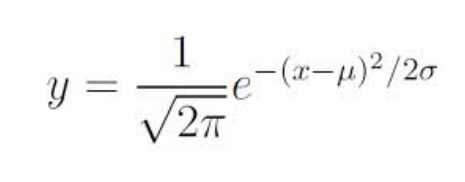
Çeyrek Sapma: (Q3 - Q1)/2

**Değişim Katsayısı (Coefficient of Variation):**

Standart sapmanın aritmetik ortalamaya bölünüp yüzle çarpılması ile elde edilen katsayıdır. Oransal değişim ölçüsü olarak dağılımlardaki ölçek etkisini gidermiş olur. Bu nedenle farklı örneklerin karşılaştırılmasıda kullanılır. Değişim katsayıları karşılaştırıldığında, değişim katsayısı küçük olan serilerde dağılımın aritmetik ortalama etrafında daha yoğun dağıldığı anlaşılır.

**Ders05 - Normal Dağılım**

Normal dağılım; 'normal dağılım eğrisi (normaly distribution curve)' ile kavramlaştırılan hipotetik bir evren dağılımıdır. 'Gauss dağılımı' ya da 'Gauss eğrisi' olarak da bilinen normal dağılım eğrisi, sürekli ve olasılıklı bir fonksiyon eğrisidir. Normal dağılım eğrisinin fonksiyonu aşağıdaki şekildedir:



Formülden de anlaşılacağı üzere normal dağılımı tanımlayan parametreler evren ortalaması (µ) ve evren standart sapması (σ)'dır. Normal dağılım eğrisi, bir fonksiyon olarak kartezyen düzlemde yandaki gibi çizilir:

Normal dağılım eğrisi, ölçme sonuçlarının orta noktalarda yoğunlaştığı, uç noktalarda seyrekleştiği bir dağılımın şeklini ifade etmektedir.

Normal dağılımın karakteristik özellikleri şu şekildedir:

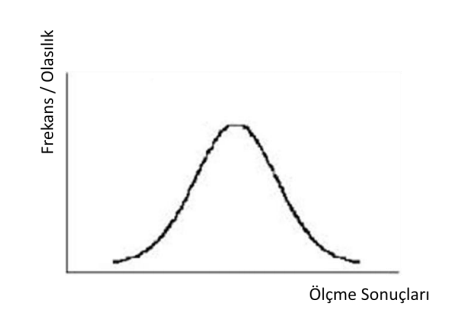
1. Simetriktir.

2. Asimptotiktir.

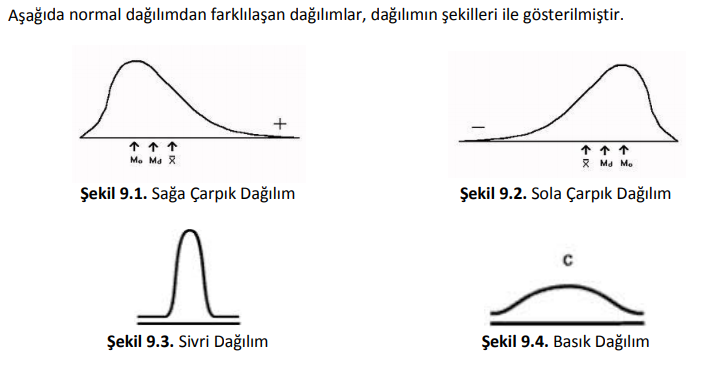
3. (-∞, +∞) aralığında değerler alır.

4. Eğri altındaki toplam alanın olasılığı 1'dir. [ P (-∞ < X < +∞) = 1 ]

5. Ortalama, mod ve medyan değerleri çakışıktır. [ µ = Medyan = Mod ]

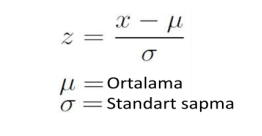


**Çarpık ve Basık Dağılımlar:**



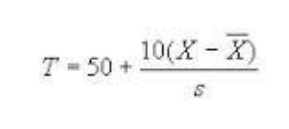
**Standart Z Puanları:**

Standart Z puanları ya da kısaca Z puanları; ortalaması 0, standart sapması 1 olan ve evrende normal dağılım gösteren 'iyi' tanımlı tipik puanlardır. Z puan dönüşümü ise ham verilerin Z puanlarına dönüştürülmesinde kullanılan doğrusal bir puan dönüşümüdür. Bir grup verinin ya da puanın ortalaması ve standart sapması bilindiğinde bu puanların her bir ayrı ayrı Z puanına dönüştürülebilir. Z puanları dönüşümünde yandaki formül kullanılır:



**Standart T Puanları:**

Standart T puanları ya da kısaca T puanları; ortalaması 50, standart sapması 10 olan ve normal dağılım gösteren 'iyi' tanımlı tipik puanlardır. T puan dönüşümü ise ham verilerin T puanlarına dönüştürülmesinde kullanılan doğrusal bir puan dönüşümüdür. Bir grup verinin ya da puanın ortalaması ve standart sapması bilindiğinde bu puanların her bir ayrı ayrı T puanına dönüştürülebilir. T puanları dönüşümünde aşağıdaki formül kullanılır:



Formülde görüldüğü gibi T puanları, Z puanlarından farklı olarak örneklem ortalaması ve örneklem standart sapması ile tanımlanmaktadır. Bunun nedenlerinden biri T puanlarının, genellikle küçük örneklemlere yönelik istatistiksel işlem ve analizlerde kullanılmasıdır.

Formül incelendiğinde Z puanları ile T puanları arasında doğrusal bir ilişki olduğu görülmektedir. T puan dönüşümü formülünde yer alan her bir değerin ortalamadan farkının alınıp standart sapmaya bölünmesi, Z puan dönüşümü formülü ile örtüşmektedir. Buna göre T puanları, Z puanlarının 10 katının 50 fazlasıdır. Yani T puanları, Z puanları kullanılarak da elde edilebilmektedir.

T = 10. Z + 50

Z puanlarının negatif değerler alabiliyor olması, bir güçlük olarak yukarıda açıklanmıştı. T puanları, Z puanlarında yaşanan bu güçlüğü gidermektedir.

Z puanlarında olduğu gibi T puanlarının da anlaşılması ve yorumlanması zordur. Fakat 'iyi' tanımlı ve tipik puanlar olmaları, uygun durumlarda daha ileri düzey istatistiksel kestirimlerde kullanılabilmesini sağlamaktadır. Bu da ilgilenilen özelliğin anlaşılması, açıklanması ve kontrol edilmesinde önemli bir hareket alanı sağlamaktadır.

**Z Dağılımı:**

Z dağılımı; ortalaması µ=0 ve standart sapması σ=1 olan Z puanlarının evren dağılımı olarak tanımlanabilmektedir. Z dağılımı olasılıklı bir normal dağılımdır. Yani Z dağılımının genel karakteristiği normal dağılımla aynıdır. Bu nedenle Z dağılımı 'standart normal dağılım' olarak da isimlendirilmektedir.

Standart Z dağılımı olarak da bilinen Z dağılımının özellikleri şu şekildedir:

1. Simetriktir.

2. Asimptotiktir.

3. (-∞, +∞) aralığında değerler alır.

4. Eğri altındaki toplam alanın olasılığı 1'dir. [ P (-∞ < X < +∞) = 1 ]

5. Ortalama, mod ve medyan değerleri çakışıktır. [ µ = Medyan = Mod ]

6. Olasılık yoğunluk fonksiyonu vardır.

Z dağılım eğrisinin olasılık yoğunluk fonksiyonu, normal dağılım eğrisinin olasılık yoğunluk fonksiyonu kullanılarak elde edilebilmektedir. Normal dağılım eğrisinin olasılık yoğunluk fonksiyonu aşağıdaki gibidir.

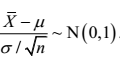
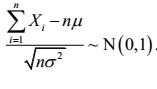


Bir Z dağılımı üzerinde belli aralıklarda eğri altında kalan alanın yani belli değerlerin görülme olasılığının belirlenmesinde temelde üç işlem yapılmaktadır:

1. İstenen alanı belirleme ve dağılım üzerinde gösterme
2. Z puanlarına dönüştürme
3. Z değerleri tablosunu okuyarak istenen alanı belirleme

**Merkezi Limit Teoremi:**

µ ortalamasına ve σ standart sapmasına sahip bir ana kütleden n büyüklüğünde rastgele tekrarlamalı örnekler alınırsa, n’in yeterince büyük olması halinde, örnek ortalamalarının dağılımı, ortalaması ve ana kütlenin ortalaması µ’ ye yaklaşık olarak eşit ve standart sapması ana kütlenin standart sapması σ bölü karekök n’ e yaklaşık olarak eşit bir normal dağılım olacaktır. n sayısı arttıkça yaklaşım daha doğru sonuç verecektir.

**Ders06 - Örnekleme Dağılımları (Sample Distributions)**

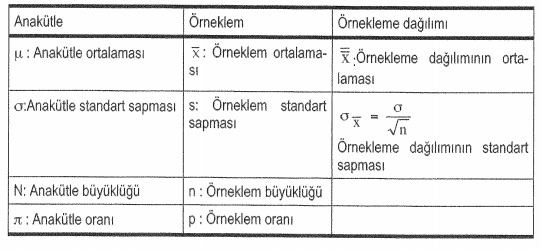
İstatistik dağılımları, istatistiğin değerinin hesaplanacağı örneğin alınmış olduğu ana kütlenin dağılımına bağlıdır. X1, X2, X3,… Xn bir değişken vektörü olarak örneklemi ifade eder. Bu örneklemdeki her bir değişkenin dağılımı ana kütlenin dağılımı ile aynıdır. Bu örneklemin gerçekleşmiş hali x1, x2, x3,… xn örneğidir. Örneklemdeki değişkenlerin bir fonksiyonu olan değişken bir istatistik’ tir.

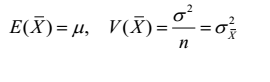
Bir istatistiğin değeri, anakütleden çekilecek belirli bir örnekleme bağlıdır. Bu nedenle, örneklem istatistiği bir rassal değişkendir ve bir olasılık dağılımına sahiptir. İstatistiğin olasılık dağılımına örneklem dağılımı (sampling distribution) adı verilir.

Örnekleme dağılımları bir ana kütleden çekilebilecek tüm olası örneklemlerden yararlanılarak oluşturulan teorik dağılımlardır. Bu dağılımların da diğer dağılımlar gibi ortalaması ve standart sapması söz konusudur.

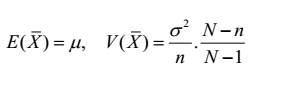
-Ortalamaların örnekleme dağılımı (Sampling distribution of the mean)

-Oranların örnekleme dağılımları (sampling distribution of the proportion)

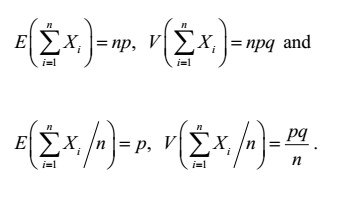




İstatistiğin örneklem dağılımının ortalaması, anakütle ortalamasına eşit, µ olup varyansı;



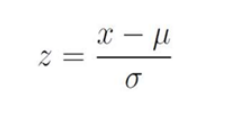
**Bernoulli Dağılımı;** Dağılımının ortalaması ve varyansı ise aşağıdaki formüllerle hesaplanır.



Student T Dağılımı:

'Student t dağılımı' ya da kısaca 't dağılımı'; normal dağılım ve Z dağılımının da içerisinde bulunduğu 'sürekli olasılık dağılımları' ailesinde yer alan dağılımlardan bir diğeridir. T dağılımı, William Sealy Gosset'in 1908 yılında Biometrika dergisinde 'Student' takma adıyla yayımladığı makalesinde tanımladığı bir hipotetik dağılımdır. T dağılımı, pratikte, Z puanları üzerinde bir düzeltme ile elde edilen t puanlarının dağılımıdır.

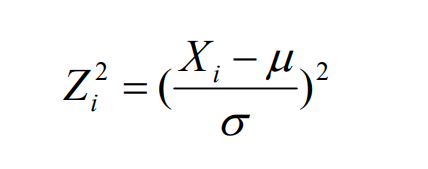
Bilindiği gibi Z puanları, evren ortalaması ve evren standart sapmasına göre tanımlanmaktadır:



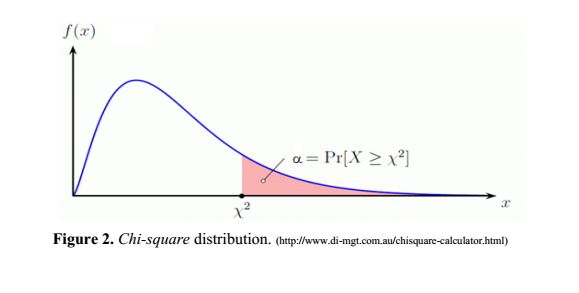
Yine bilindiği gibi evren ve örneklem standart sapması farklı formüllerle ve farklı şekillerde hesaplanmaktadır. Bu farklılık, gözlem sayıları ile ilişkilidir. Gözlem sayısı arttıkça yani örneklem büyüklüğü evrene yaklaştıkça, örneklem değerleri yani istatistikler, evren değerlere yani parametrelere yaklaşmaktadır. Diğer bir deyişle küçük örneklemlerde yapılan kestirimlerin evren değerlerden yani parametrelerden farklılaşma olasılığı daha yüksektir. Bu nedenle örneklem üzerinde ve özellikle küçük örneklemler üzerinde yapılan kestirimlerde, daha 'sağlam' kestirimler elde edilmesi için istatistiksel düzeltmeler yapılır. Örneklem standart sapmasının hesaplanmasında (n-1)'e bölme bu tür bir düzeltmedir. İşte Gosset, Z dağılımının temelde bir evren dağılımı olduğu, bu dağılımın küçük örneklemlerde kullanılabilmesi için bir düzeltme yapılması gerektiği fikrinden hareketle 'Student t Dağılımı' olarak bilinen dağılımı geliştirmiştir. Bu düzeltmeyi, Z dağılımının olasılık yoğunluk fonksiyonundan hareketle geliştirdiği t dağılımına özgü bir olasılık yoğunluk fonksiyonu tanımlayarak göstermiştir. Buna göre t dağılımının olasılık yoğunluk fonksiyonu aşağıdaki gibidir:

**Ki-Kare Testi (Chi-Square):**

Örnekleme yoluyla elde edilen rakamların, anakütle rakamlarına uygun olup olmadığı; bir başka ifadeyle gözlenen değerlerin teorik (beklenen) değerlere uygunluk gösterip göstermediği ki-kare testi ile tespit edilir. Standart normal değişken Z değerlerinin;



şeklinde karesi alındığında, Zi^2 değerlerinin dağılımı ki-kare dağılımına dönüşür.



**F Dağılımı:**

F Dağılımı, Sir Ronald Aylmer Fischer tarafından bulunmuştur. Temelde Fischer'in F dağılımı, beklenen ve gözlenen değerlerin uyumunu test etmeye yarayan chi-square (kay kare/ki kare) dağılımından türetilmiş bir dağılımdır.



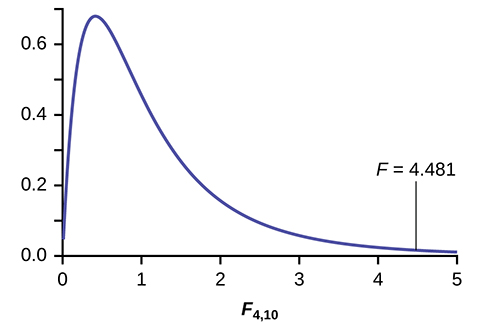


Burada pay ve paydadaki ki kare değerleri birer tesadüfi değişkenlerdir.

Dolayısı ile bunların oranı olan F de bir tesadüfi değişkendir.

F değeri sadece pozitif değerler alır.

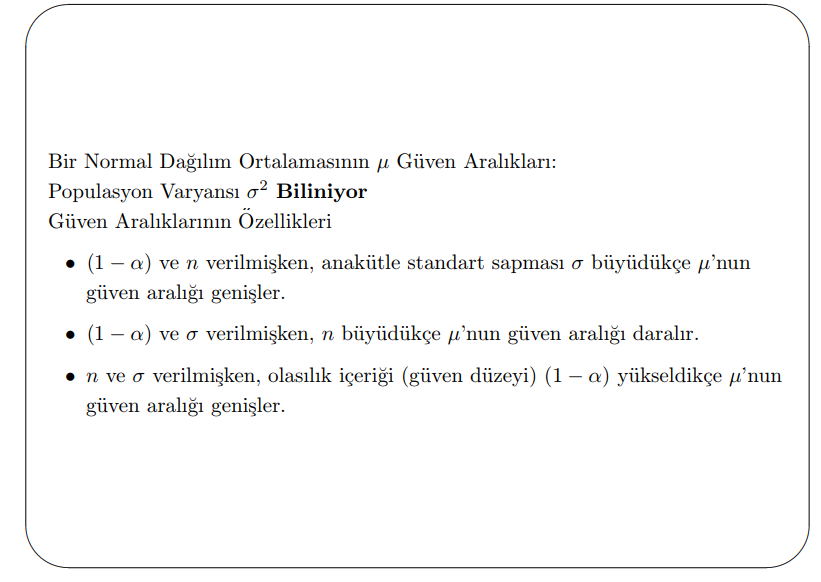
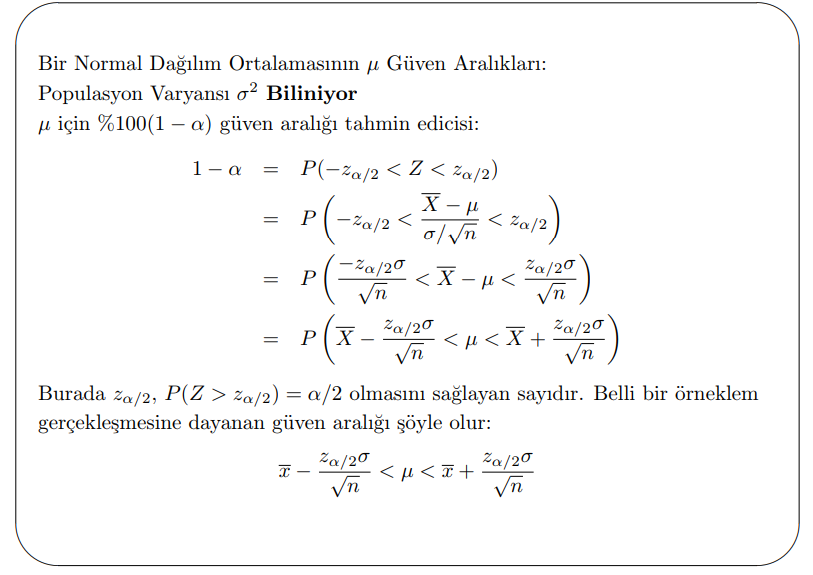
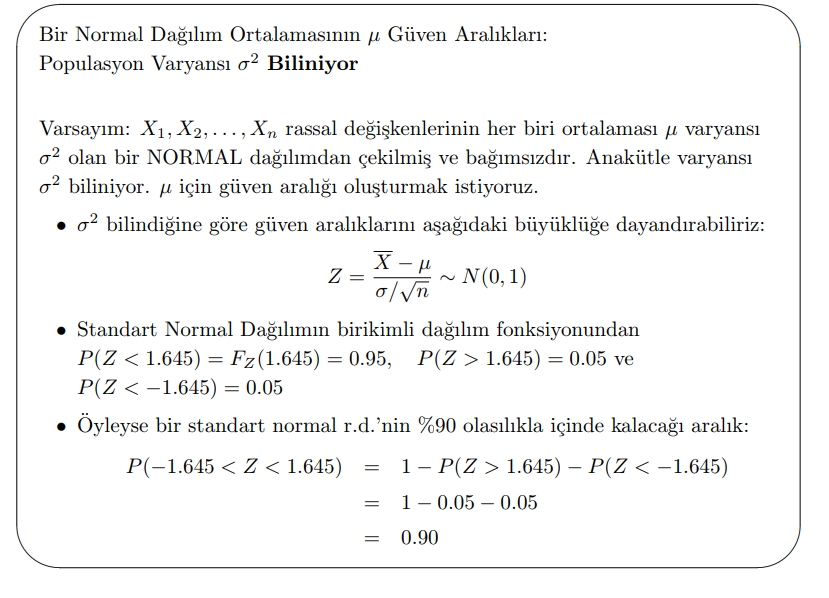
F, ki kareden farklı bir değişkendir ve farklı bir olasılık dağılımına sahiptir.



**Ders07 - Aralık Tahmini (Interval Estimation)**

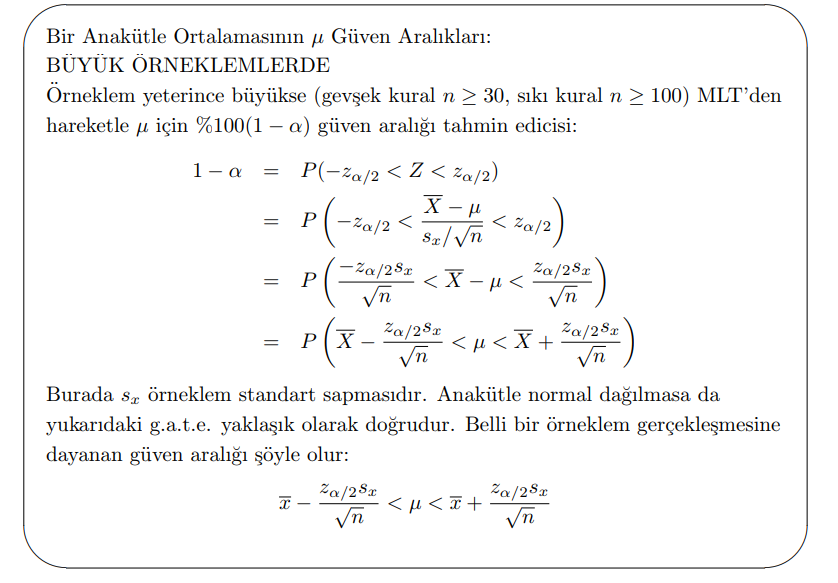
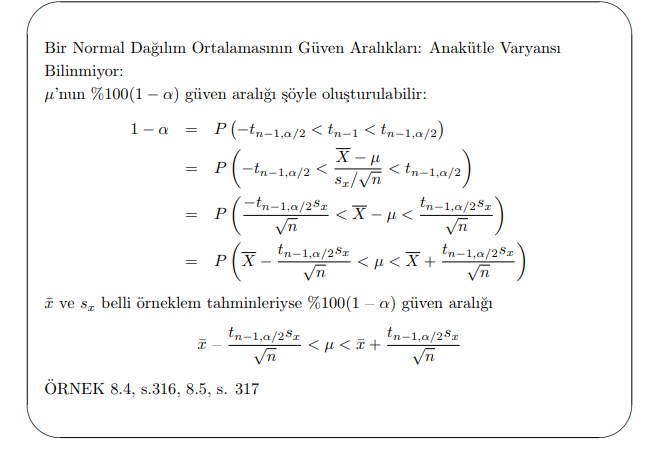
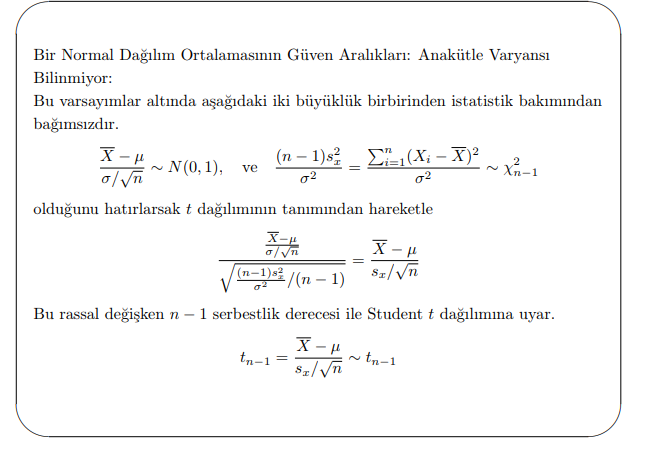
* Aralık tahmini bilinmeyen parametreye ilişkin belirsizliği açık olarak yansıtır.
* Bir anakütle katsayısının aralık tahmin edicisi, o anakütle katsayısının içine düşebileceği bir aralığı örneklem bilgisine dayanarak belirlemenin kuralıdır.
* Buna karşılık gelen tahmine de “aralık tahmini” denir.
* θ bilinmeyen popülasyon parametresi olsun. θ’nın aralık tahmin edicisi P(A < θ < B) = 1 − α, B > A
* Burada A ve B örneklem bilgisine bağlıdır ve rassal değişkendir. A: alt güven sınırı, B: üst güven sınırı
* 1 − α: güven düzeyi ya da olasılık içeriği
* A ve B’nin belli örneklem gerçekleşmelerini a ve b ile gösterirsek, bu (a, b) aralığına θ’nın %100(1 − α) güven aralığı denir: a < θ < b

**Varyans Biliniyor İse:**



**Varyans Bilinmiyor İse:**

**n>30 n<30**

****

**Ders08 - Hipotez Testi (ÖZET OLDUĞUNU UNUTMAYIN KARIŞIK GELEBiLiR !!!)**

**Hipotez Nedir ?:**

Örnekleme ile test edilmeye çalışılan bir popülasyonun ilgili parametresi hakkında ortaya sunulan iddiadır.

• Örneğin;

– A dersi için vize ortalaması 50’nin altındadır.

– Firestone ve Lassa firmalarının ürettikleri lastiklerin kaliteleri aynıdır.

• Görüldüğü gibi bir konu hakkında öne sürülen ve doğruluğu henüz ispatlanmamış görüşler hipotezlerdir. Hipotezler üzerinde çeşitli işlemler yapılarak ifadenin “doğruluğu/yanlışlığı” araştırılır.

**Hipotez Testi ve Aşamaları:**

Popülasyonu incelemeye yönelik yapılan çalışmalar ve bunların raporlanması ile hipotezin kabul edilip edilmeyeceğinin belirlenmesi işlemine hipotez testi denir. Hipotez testi aslında bir nevi karşılaştırma ve seçim işlemi olduğu için birden fazla hipoteze ihtiyaç duyulur. Bu hipotezlere ise **alternatif hipotez** denir.

1. **NULL ve ALTERNATİF HİPOTEZLERİN BELİRLENMESİ:**

Popülasyon parametresine genellikle belli bir değer atanır ve bu öne sürülen temel iddia null hipotezidir. Null hipotezi sıfır/başlangıç hipotezi olarak da bilinir. H0 ile gösterilir. Mevcut veriler null hipotezinin doğruluğu hakkında şüphe uyandırdığında kıyas yapmak için ortaya sunulan ikinci hipotez alternatif hipotezdir. Yapılan işlemler eğer H0 ’ı yanlış çıkarırsa bu H1 ’in kabulü anlamına gelir.

1. **ÖNEM veya RİSK DERECESİNİN BELİRLENMESİ:**

Genellikle risk derecesi olarak %5=0,05 ve %1=0,01 kullanılmakla birlikte bu tercihi bir durumdur. Risk derecesi temelde doğru olan null hipotezinin reddedilme olasılığını gösterir. Risk derecesini belirleyerek hipotez testi sırasında yapılabilecek hataları minimuma indirmek isteriz. Bir hipotez testi sırasında null hipotezinin doğruluk/yanlışlık ve kabul/reddedilme durumlarına göre 2 tip hata yapılabilir.(1.tip ve 2.tip hata)

!!! Null hipotezi doğru iken reddedilirse 1. tip hata, yanlış iken kabul edilirse 2.tip hata yapılmış olur.

3. **İSTATİSTİKSEL TEST METODUNUN BELİRLENMESİ:**

Örneğin F,t,ki kare istatistiksel testleri kullanılarak null hipotezi ile ilgili değerin bulunması işlemidir.

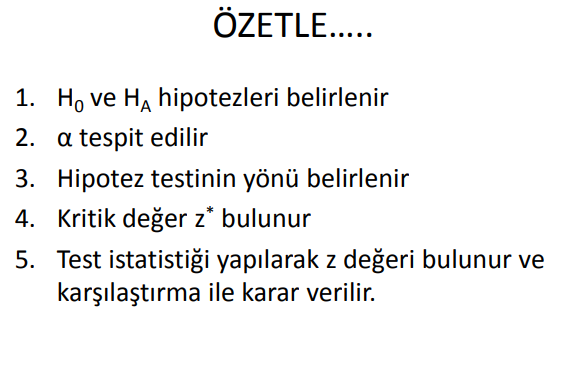
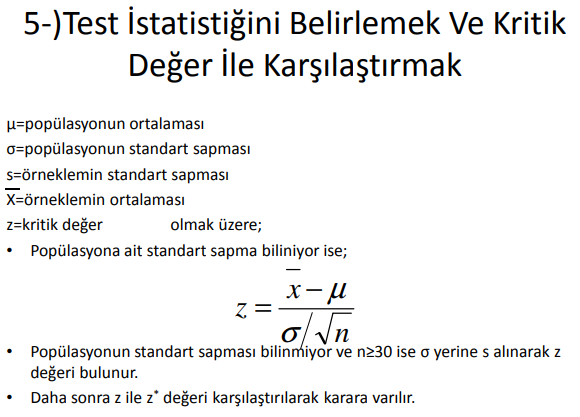
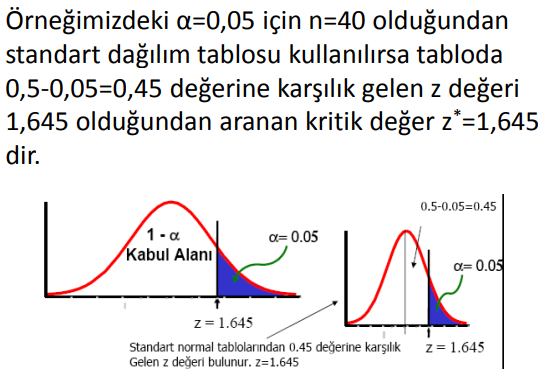
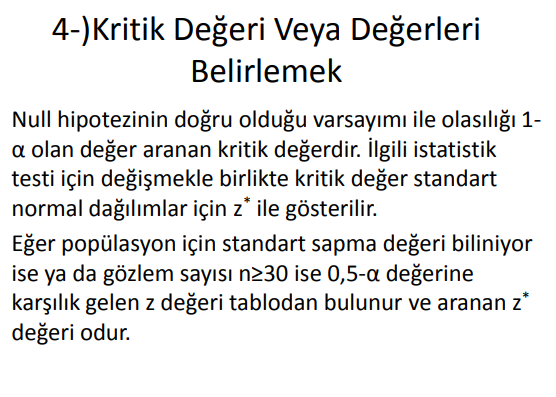
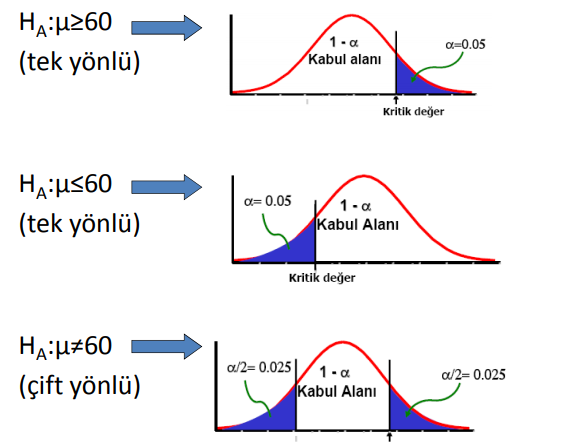
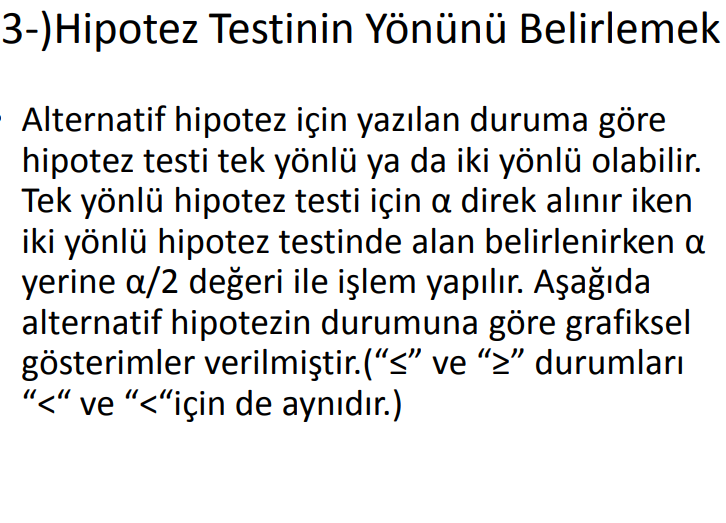
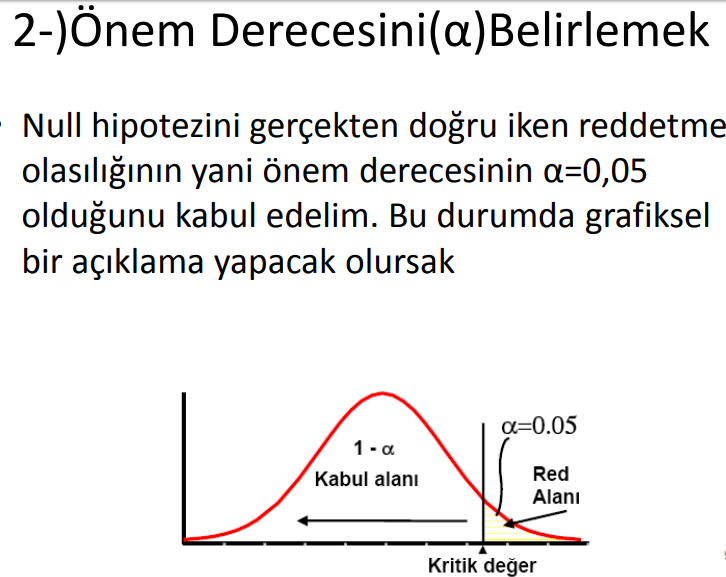
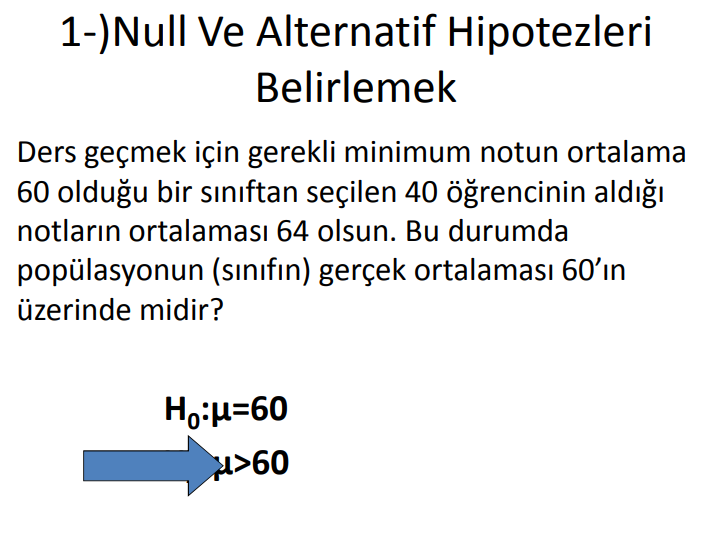
1. **NULL HİPOTEZİNİN KABUL/RED DURUMUNUN BELİRLENMESİ:**

Yukarıdaki maddede (3) bulunacak değerin durumuna göre null hipotezinin kabul/red koşullarının belirlenmesidir.

1. **NULL HİPOTEZİ İÇİN KARAR VERME:**

Yapılan işlem sonuçlarına göre null hipotezinin kabul edilip edilmeyeceği belirlenir.

**EĞER n>=30 İSE;**

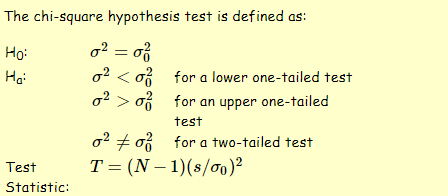


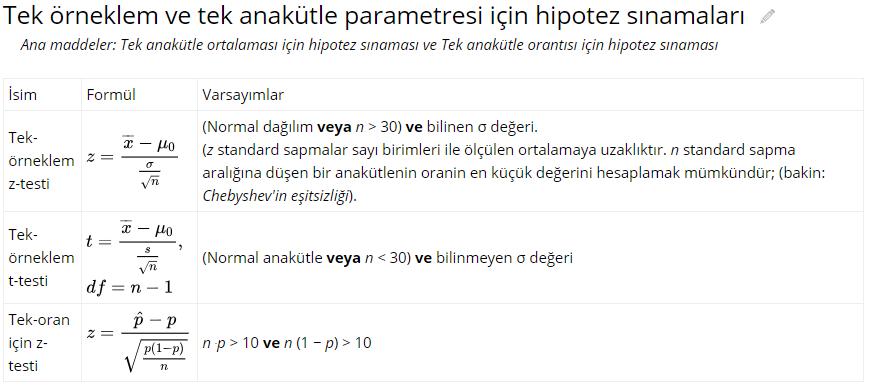
**EĞER n<30 İSE;**

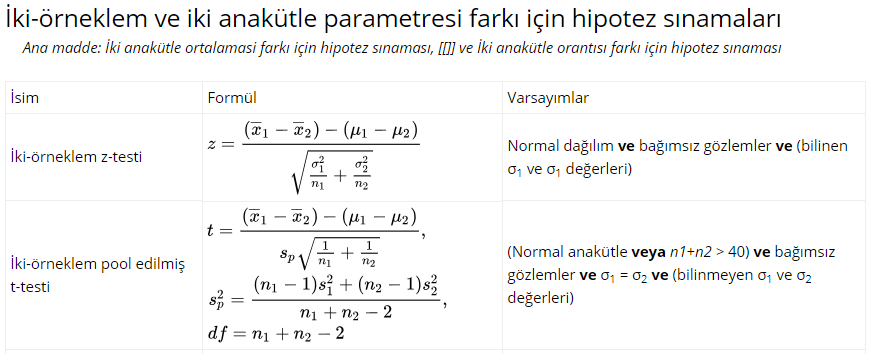
!!!Tüm her şey aynı sadece z değeri yerine t değeri hesaplanır.

**EĞER BİLDİĞİMİZ TEK ŞEY VARYANS VE STANDART SAPMA İSE;**

Ki - Kare Kullanılır. Yani yandaki işlem --->>

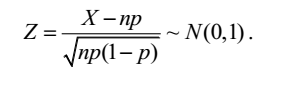




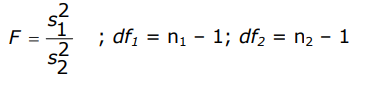


**Tek Ana Kütle Oranı İle İlgili Hipotez Testi:**

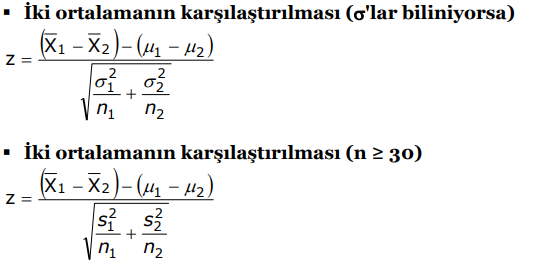
Ana kütlenin herhangi bir niteliğinin belirli bir orandan büyük, küçük veya farklı olup olmadığının test edilmesinde kullanılır.



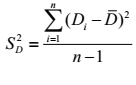
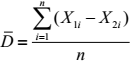
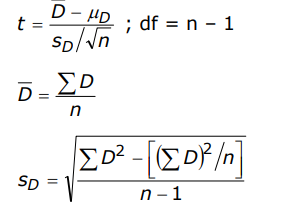
**Varyans Eşitliği / Homojenliği İçin F Testi:**



**Bağımsız İki Örneklemde Ortalamaların Karşılaştırılması:**



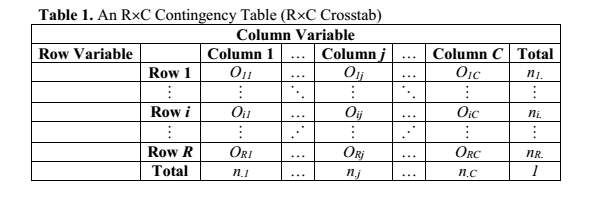
**Bağımlı / Eşleştirilmiş İki Ortalama Arasındaki Hipotez Testi:**

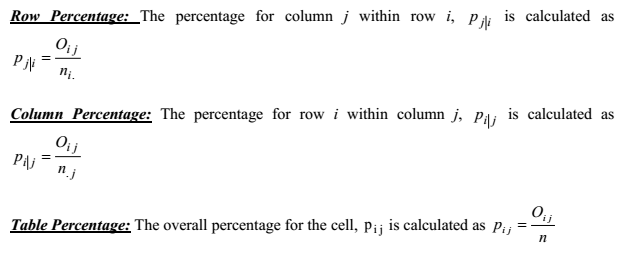


**\*\* DİĞER TÜM FORMÜLLERDE OLDUĞU GİBİ BU FORMÜLDE DE DİKKATİMİZİ CELBEDEN BİR ŞEY VAR Kİ, O DA N>=30 İSE “Z DAĞILIMI”,**

**N<30 İSE “T DAĞILIMI” KULLANILDIĞI !!!**

**Ders09 - Olasılık Tablosunda Ki-Kare Testi**

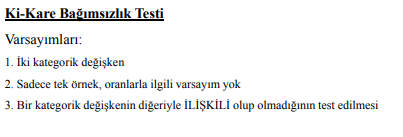
 



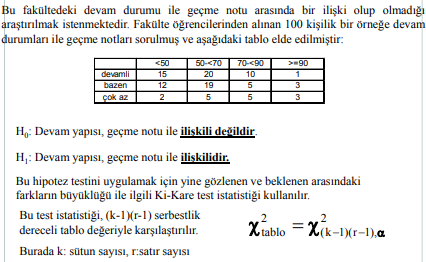
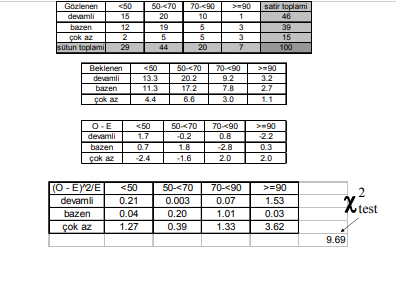
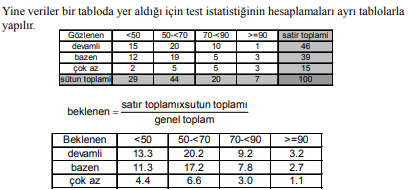
**Pearson’s Ki-Kare Testi:**

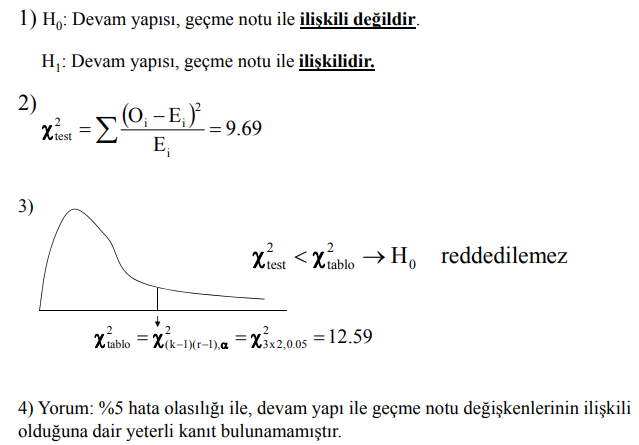
Sık sık karşılaşılan araştırma sorularından biri de iki değişkenin birbiri ile ilişkili olup olmadığıdır. Mesela, bir sosyolog okul vasıtasıyla kazanılan eğitim seviyesinin gelirle ilişkili olup olmadığını araştırmak isteyebilir. Bir okulda çalışan beslenme uzmanı öğrencilerin beslenme uzmanı öğrencilerin beslenme seviyelerinin akademik performanslarıyla ilişkili olup olmadığını bilmek isteyebilir.

İki değişken arasında ilişki yoksa, birisinin dağılımı hiçbir şekilde diğerinin dağılımına bağımlı değilse bu iki değişken birbirinden bağımsızdır denir. İki değişken arasında ilişki yoksa, belirli bir değişkenin değerini bilmek, diğer değişkenin değerini tespite imkan sağlamaz.



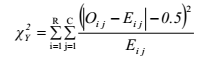
Örnek üzerinde anlatmak kolay olduğundan; **O: Gözlenen, E: Beklenen**

****

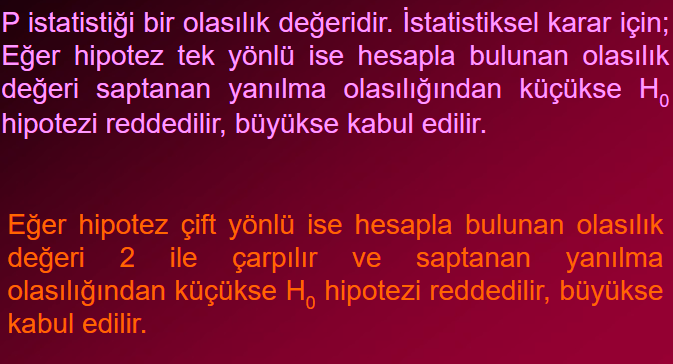
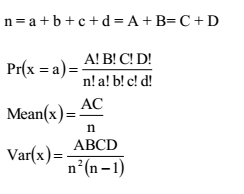
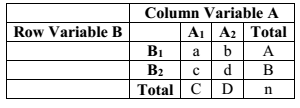
****

**Yate’s Ki-Kare Testi:**

2x2 lik Tablolarda Uygulanır.



**Fisher Ki-Kare Testi:**



**\*\* CHI-SQUARE TEST FOR CONTINGENCY TABLES (SON 3 SAYFA) & INTERVAL ESTIMATION FOR TWO POPULATIONS YOKTUR !!!**