I. İSTATİSTİK VE OLASILIK

Doç. Dr. İrfan Yolcubal
Kocaeli Üniversitesi
Jeoloji Müh. Bölümü

Ders Kitabı

- Statistical analysis of Geological data (Koch G. S., ve Link, R. F., 1980. Dover Publications) A data-based approach to statistics (Iman, R. L., 1994)
- Basic statistics for Business and Economics (Lind, D. A., and Mason, R. D., 1997)
- İstatistik Analiz Metotları (Prof. Dr. Bilge Aloba)

DEĞERLENDİRME

- Devam zorunlu (% 70)
- 2 Sınav (Ara vize + Final)
- Ödev
- Grup çalışması OK
- Ödev kopyalamak yasak

DERS PROGRAMI

- Data toplama ve sunum şekilleri
  - Örnek vs. Popülasyon kavramları
  - Data toplama teknikleri
  - Data sunum şekilleri
    - Pasta diyagramlar
    - Histogramlar
    - Bar grafikler
    - Kümülatif rölatif sıklık grafikleri
    - Dağılım grafikleri (X-Y)
- Dataların Değerlendirilmesi
  - Tarifsel istatistik
    - Analitik ve analitik olmayan ortalamalar
    - standart sapma, varyans, standart hata, güvenilirlik aralığı vb…

DERS PROGRAMI devam

- Olasılık ve Olasılık dağılımları (Probability density functions)
  - Binom
  - Logaritmik
  - Normal
  - Poisson
- Tahmin ve Hipotez testi
  - t-test
  - z-testi
  - Varyans analizi (ANOVA)
- Korelasyon ve Regresyon
- Linear regresyon
- Jeolojide örnekleme ve jeolojik datalardaki değişkenlik

İSTATİSTİK NEDİR?

- İstatistik kısaca, data analizini kapsayan matematik biliminin alt bir dalıdır.
  - Dataların toplanması, derlenmesi, özetlenmesi,sunumu, analizi ve aynı zamanda verilerden geçerli bir sonuç çıkarılması istatistik dalının başlıca ilgi alanlarıdır.

İSTATİSTİĞİN UYGULAMA ALANI

- İstatistiksel metotlar, Jeoloji'de olduğu gibi, modern yaşamın büyük bir alanında da, dataların değerlendirilmesinde ve analizinde yaygın bir şekilde kullanılmaktadır.
  - Japon ürünlerini dünyada popüler yapan kalite-kontrol tekniklerinin uygulanmasında
  - Ozon tabakasındaki incelmenin tahmininde
  - Nüfus sayımında
  - TOP 40 hit listesinin belirlenmesinde
  - Hava tahminlerinde
  - TV reytinglerin belirlenmesinde
  - Kişisel bilgisayarınızdaki parçaların performanslarının geliştirilmesinde
  - Seçim tahminlerinde
  - Risk analizinde
  - Ve daha bir çok alanda……

İSTATİSTİK TÜRLERİ

- Tarifsel (Descriptive) istatistik: nümerik verileri derlemek, düzenlemek, ve özetlemek için kullanılan prosedürler
- Tümevarımsal (Inferential) istatistik: örneklemeye dayanarak bir popülasyon hakkında bilgi elde etmek için kullanılan metotlar

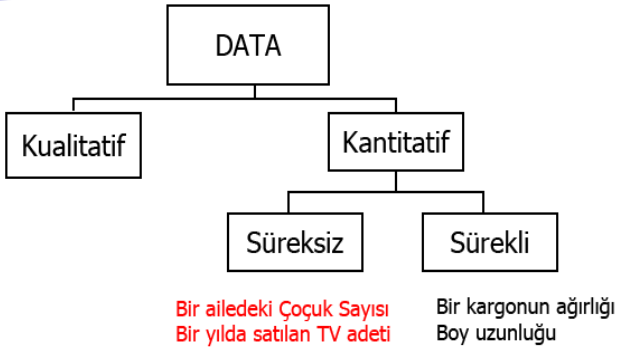
DATALARIN TOPLANMASI: Kavramlar

- Örnek (Sample) : İnceleme yapılan popülasyonun bir bölümü
- Popülasyon: Hakkında bilgi edinilmeye çalışılan birimlerin (kişiler, nesneler, deneysel sonuçlar, vb. gibi) toplamını oluşturmakta.

Değişken Türleri

- Kantitatif (Quantitative) Değişken:
  - Sayısal ölçekte ifade edilir.
  - Miktar hakkında bilgi verir.
  - Hesabınızdaki bakiye, pilin ömrü, sınıftaki öğrencilerin sayısı vb.
- Kualitatif (Qualitative)Değişken :
  - Nümerik olmayan değişkenlerdir.
  - Doğum yeri, göz rengi, ırk, vb…

Değişkenlerin Sınıflandırılması



DATA TOPLANMASI

- Örneklemeye geçmeden önce inceleme yapılan popülasyonun iyi bir şekilde belirlenmesi gerekmektedir.
- Uygun örnekleme tekniği ve protokolü: Toplanan örneklerin incelenen popülasyonu tam anlamıyla yansıtması gerekmektedir.

DATA TOPLANMASI

- Tüm popülasyonu incelemek (Sayım) yerine neden örneklemeyi kullanıyoruz?
  - Düşük maliyet
  - Zaman
  - Dikkatlice alınmış örnekler bazı durumlarda bir sayım'dan daha doğru bilgi verebilir.
  - Bazı durumlarda imkansız olabilir.
  - Ürünlerin yok edilerek test edilişinden örnekleme tek başına yeterli olabilir.

ÖRNEKLEME PLANININ AMAÇI

- Yüksek kalite: Toplanan verilerin doğruluk derecesi
- Savunabilirlik: Planın geçerliliği ile ilgili dokümantasyonun mevcut olması
- Tekrarlanabilirlik: Örnekleme planını takip ederek verilerin tekrar üretilebilmesi
- Temsil edici olması: İncelenen popülasyonu tamamıyla temsil etmesi
- Faydalı olması: Toplanan veriler planın amacına ulaşmasında kullanılabilir olması

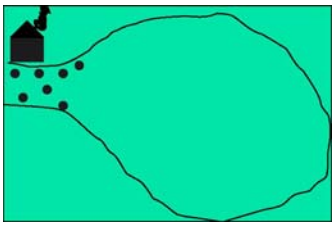
ÖRNEKLEME PLANI

- Örnekleme planının hazırlanışında örneklerin nerede ve ne zaman alınacağına karar vermek gerekmektedir.
- Örneklerin sayısı, lokasyonu, ve zamanı örnekleme bütçesini aşmadan istatistiksel olarak geçerli bir örnek almaya yeterli olmalıdır.
- Bunu sağlamak için uygun bir örnekleme stratejisi belirlemek gerekmektedir.

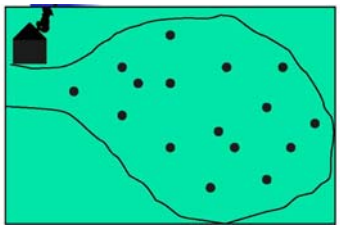
ÖRNEKLEME ŞEKİLLERİ

- Rasgele (Random): Her örneğin aynı sayıdaki gözlemde eşit olasılıkla olarak seçilebilmesi
- Sistematik:
- Karara dayalı (Judgemental):

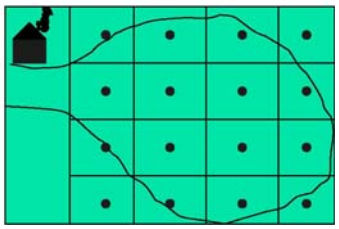
ÖRNEKLEME ŞEKİLLERİNE BİR ÖRNEK



Karara dayalı



Rasgele



Sistematik

Örnekleme Hatası (Sampling Error)

- İstatistiksel anlamda hatadan ziyade örneklerin birbirlerinden olan doğal değişkenliklerini temsil etmektedir.
- Tüm toplanan örneklerde bir tane ortak özellik bulunmakta: örneklerin hiçbirisinin tamamıyla tüm popülasyonu temsil etmemesi
- Dolayısıyla, örneklemeye dayalı tahminler ile popülasyonun gerçek karakteristiği arasında daima farklılık olacaktır.

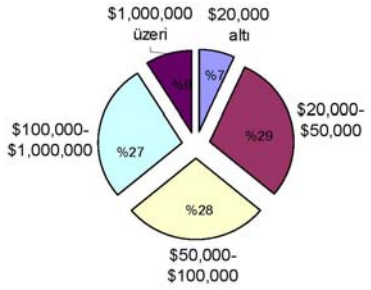
VERİ GRAFİKLEME TÜRLERİ

- Pasta diyagramlar (Pie Charts)
- Bar grafikler
- Histogramlar
- Kartezyen (X-Y) grafikleri
- Frekans Dağılım grafikleri

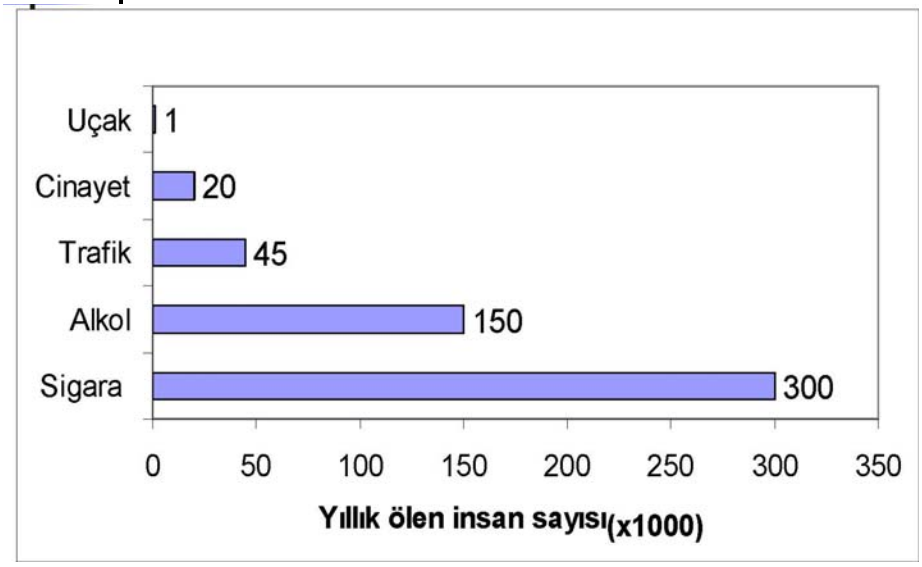
Pasta Diyagramlar

- Farklı yada kantitatif dataların oran yada yüzde şekilde sunulmasında kullanılır.

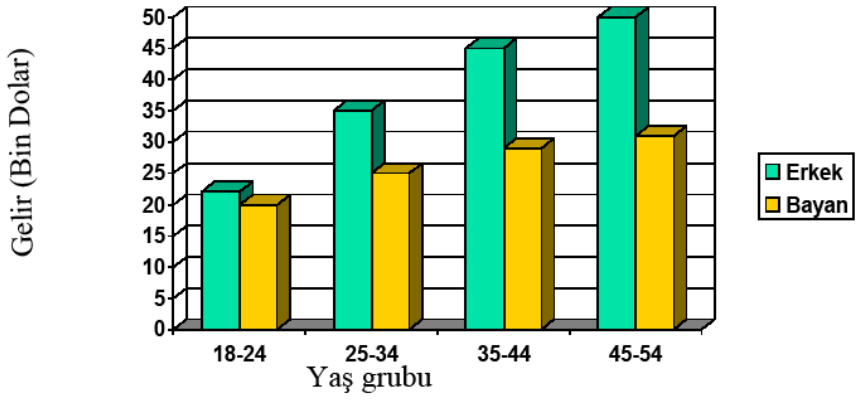
Farklı Gelir gruplarına göre ödenen yıllık vergi miktarlarının dağılımını gösteren pasta diyagramı



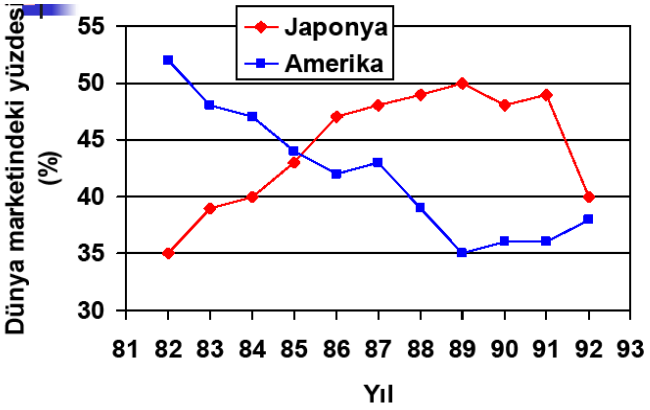
BAR GRAFİKLERİ



Düşey bar grafikleri:Üniversite mezunu erkek ve bayanların yaş gruplarına göre gelir dağılımını gösteren bar grafiklerine bir örnek



X-Y GRAFİKLERİ



II. VERİLERİN ORGANİZE EDİLMESİ VE SUNULMASI

Tasnif

- Bir kitlenin veya grubun özelliklerine göre yapısını ortaya çıkarabilmek amacıyla, elde edilen bilgileri bir vasıf veya vasıflar bakımından çeşitli şıklara ayırarak aynı şıkkı ait birimleri kümeler halinde bir araya getirme işlemine denir.
- Veri sayısının sınırlı olduğu durumlarda uygulanabilir.

Tasnif'e örnek: 100 kişilik bir sınıftaki öğrencileri yaş vasfına göre tasnif edersek

Yaş	Frekans
18	21
19	25
20	30
21	18
22	6
Toplam:	100

Gruplama

- Eğer tasnif edilecek veri sayısı çok fazla ise bunları tasnif yoluyla kümelere ayırmak mümkün olsa bile anlamlı ve işlemlere elverişli olmayabilir. Böyle durumlarda bir vasfın birbirine yakın olan şıklarını gruplar halinde toplamaya, yani gruplamaya başvurulabilir.

Gruplamaya örnek: Dünyadaki 29 en büyük şehir nüfus itibariyle gruplanarak bir frekans dağılımı veya bölünmesi şeklinde ifade edilebilir.

Nüfüs Sınıfları (1000 kişi olarak)	Şehir sayısı (Frekans)
3000-4000'den az	6
4000-5000	3
5000-6000	4
6000-7000	4
7000-8000	4
8000-9000	4
9000 ve üstü	4
Toplam	29

Gruplamaya örnek:(Hatalı) Bir endüstri dalında faaliyet gösteren işletmelerde çalıştırılan işçi sayısına göre gruplamak istersek

Çalışan sayısı	Frekans
1-2	315895
3-4	40588
5-9	9508
10-19	2348
20-49	721
50-59	44
100 ve daha fazla	68
Toplam	369133

FREKANS DAĞILIMLARI YADA BÖLÜNMELERİ (Frequency Distributions)

- Verilerin her bir sınıf aralığına düşen gözlem sayısını(frekans) gösterecek şekilde gruplandırılması işlemi.

Sınıf aralığı, Sınıf sınırları, sınıf orta noktası kavramları

- Sınıf aralığı: Sınıfın alt ve üst sınırları arasındaki fark
- Sınıf sınırları:Sınıfa ait minimum ve maksimum sınır değerleri
- Sınıf orta noktası veya noktası: Sınıfın alt ve üst sınırların ortalaması

Frekans Dağılımının oluşturulması
Örnek: Bir taşıtın yıl içindeki satış fiyatlarının organize edilmemiş hali: Ham data

$20,197	$20,372	$17,454	$20,591	$23,651	$24,453	$14,266	$15,021	$25,683
27872	16587	20169	32851	16251	17047	21285	21324	21609
25670	12546	12935	16873	22251	22277	25034	21533	24443
16889	17004	14357	17155	16688	20657	23613	17895	17203
20765	22783	23661	29277	17642	18981	21052	22799	12794
15263	33625	14399	14968	17356	18442	18722	16331	19817
16766	17633	17962	19845	23285	24896	26076	29492	15890
18740	19374	21571	22449	25337	17642	20613	21220	27655
19442	14891	17818	23237	17445	18556	18639	21296	

Minimum fiyat

Maksimum fiyat

Sınıf Sayısının Belirlenmesi

- Sınıf sayısı: k ; n:toplam veri sayısı

$$2^k \geq n$$

$\cancel{2^6 = 64}$ n=80

$2^7 = 128$

Tavsiye edilen minimum. sınıf sayısı: 7

•Genel kural olarak, frekans dağılımları oluşturulurken 5' den az ve 15' den fazla sınıf kullanılmamalı.

Sınıf Aralığının Belirlenmesi

- Tavsiye edilen sınıf aralığı= $\frac{\text{Max. değer- Min. değer}}{\text{Sınıf sayısı}}$

$$\frac{33625-12546}{8} = \$2635$$

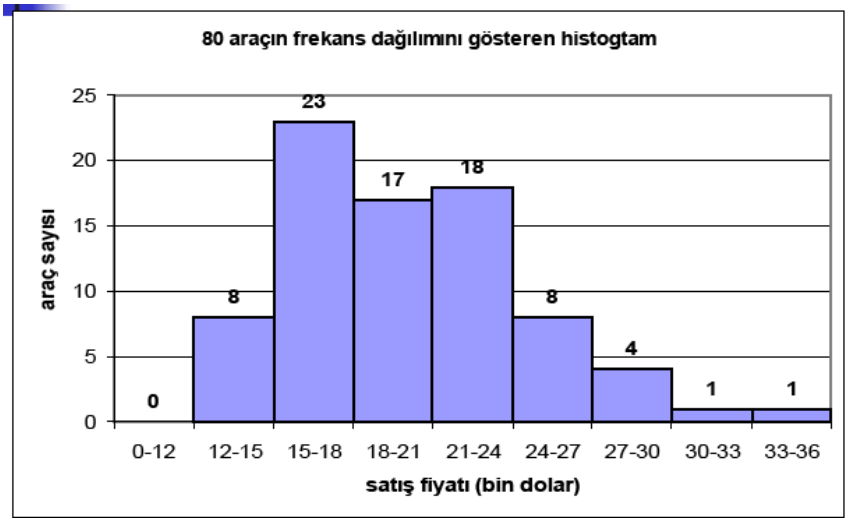
~$3000

•Sınıf aralığı seçerken yuvarlak rakamlar kullanılmalı
•Birinci sınıfın alt limiti sınıf aralığının çift katı olmalı
•Sınıf aralıkları birbirleri ile örtüşmemeli
•Açık sınıf aralıklarından sakınılmalı

FREKANS DAĞILIM TABLOSU:

Araba satış fiyatı (bin $)	Frekans
$12-15	8
15-18	23
18-21	17
21-24	18
24-27	8
27-30	4
30-33	1
33-36	1
Toplam	80

FREKANS DAĞILIM GRAFİGİ (HİSTOGRAM)



FREKANS DAĞILIMLARININ OLUŞTURULMASINDA DİKKAT EDİLEÇEK HUSUSLAR

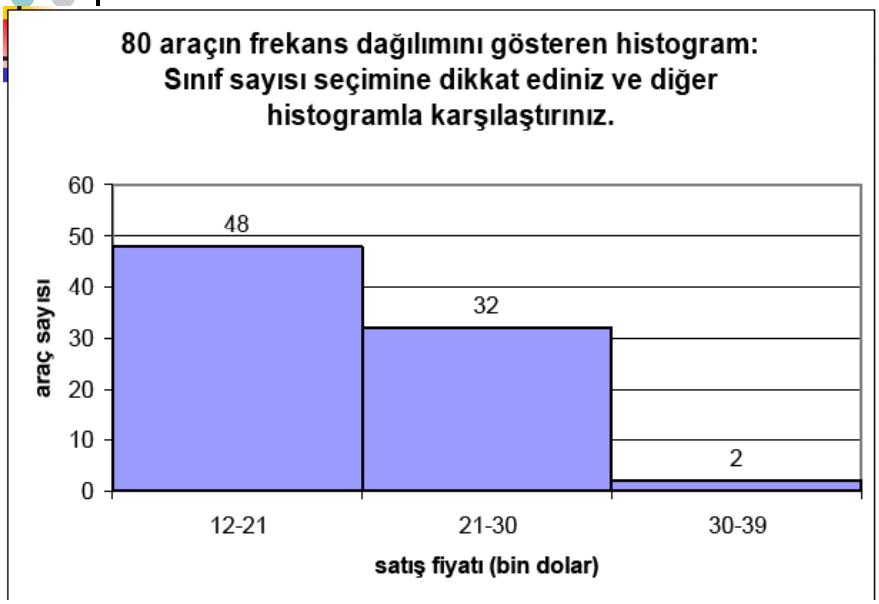
Frekans dağılımlarının oluşturulmasında mümkün olduğu kadar eşit sınıf aralıkları seçilmeli

Eşit olmayan sınıf aralıkları frekans dağılımları grafik edilirken sorun yaratabilirler.

Fakat bazı durumlarda (çok sayıda boş sınıf oluşturmaktan kaçınmak için) eşit olmayan sınıf aralıklı frekans dağılımlarının oluşturulmasında kullanılabilir.

Uygun olarak seçilmemiş sınıf sayısına göre oluşturulmuş frekans dağılımları, verinin frekans dağılımı hakkında faydalı bilgiler sunmayabilir. Örnek:

Araç satış fiyatı	Araç sayısı (Frekans)
$12000-21000	48
$21000-30000	30
$30000-39000	2
Toplam	80

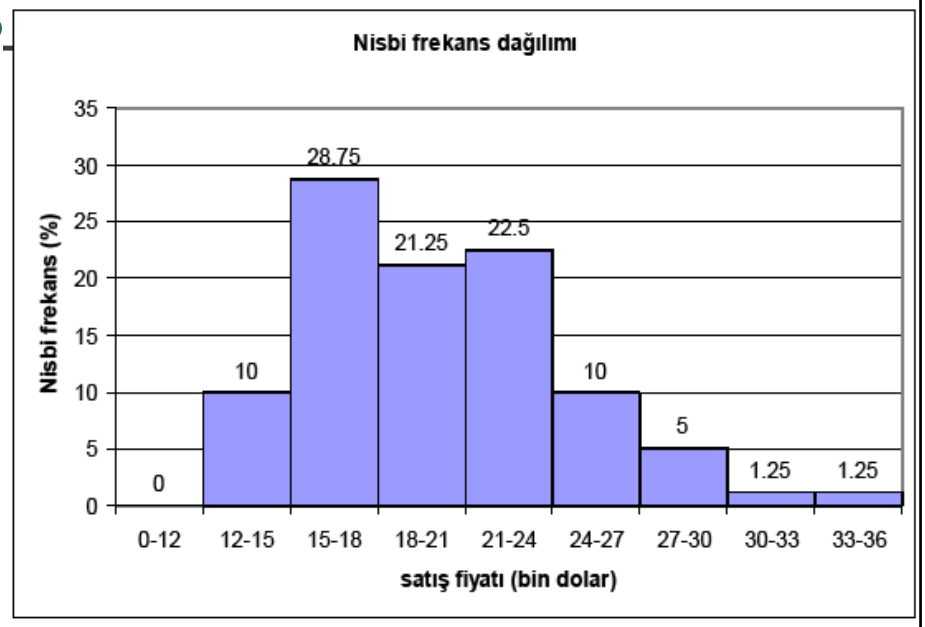


Nispi frekans dağılımları (relative frequency distributions)

- Bir çok durumda bir sınıfın mutlak frekansından çok toplam içindeki nispi frekansını bilmek gerekmektedir.
- Sınıfın nispi frekansı, o sınıfın frekansının toplam frekansa oranıdır.

Nispi frekans dağılımı

Araba satış fiyatı (bin $)	Frekans	Nisbi frekans (%)
$12-15	8	10
15-18	23	28.75
18-21	17	21.25
21-24	18	22.50
24-27	8	10
27-30	4	5
30-33	1	1.25
33-36	1	1.25
Toplam	80	100



KÜMÜLATİF FREKANS DAĞILIMLARI

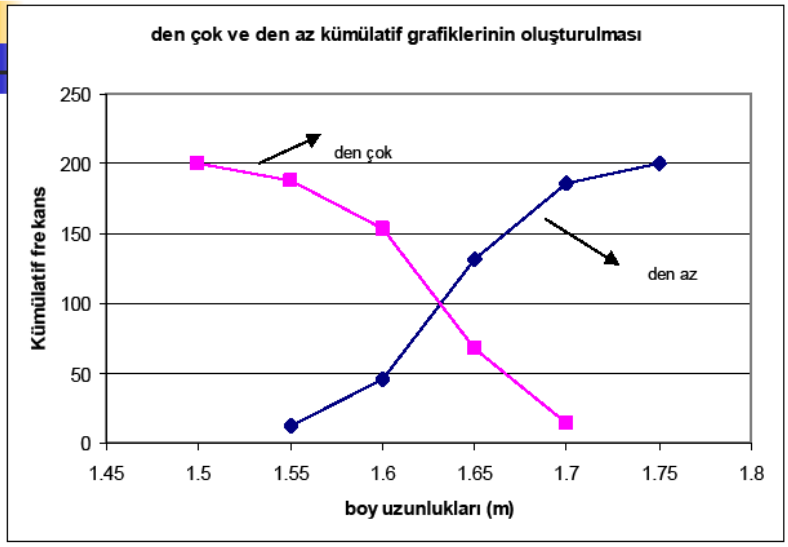
- Kümülatif frekans dağılımlarının en önemli özellikleri belirli bir düzeyin altında veya üstünde bulunan birimlerin frekansını gösterebilmeleridir.
- Kümülatif frekans dağılımları sınıf aralıkları farklı serilerin kıyaslanmasını kolaylaştırır. Kıyaslamanın sağlıklı olabilmesi için önceden mutlak frekansların nispi frekanslara çevrilmesi gerekir.

Kümülatif frekans dağılımlarının oluşturulması:
Örnek:Bir sınıftaki öğrencilerin boy uzunluklarının
frekans dağılımları

Boy uzunlukları (cm)	Frekans
150-155	12
155-160	34
160-165	86
165-170	54
170-175	14
Toplam	200

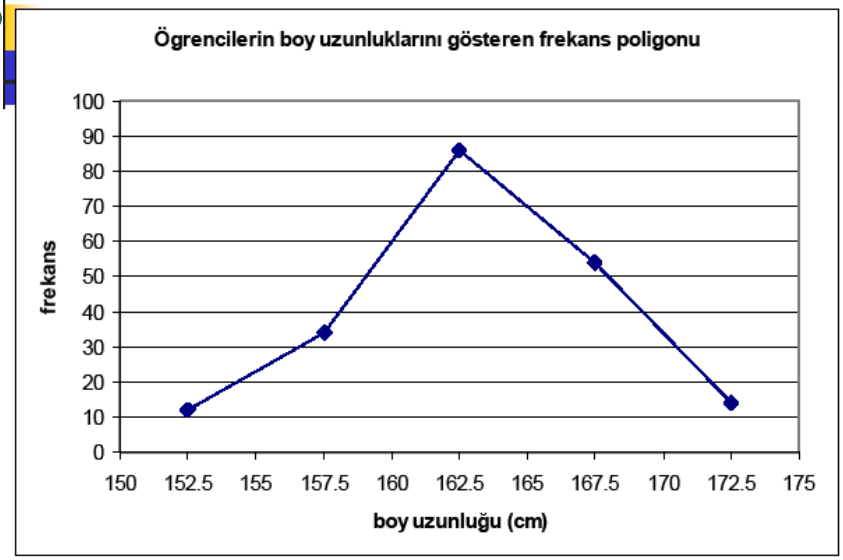
Kümülatif Frekans Dağılımları

(den az)	frekans	(den çok)	Frekans
155′ den az	12	150 ve daha çok	200
160′dan az	46	155 ve daha çok	188
165′den az	132	160 ve daha çok	154
170′den az	186	165 ve daha çok	68
175′den az	200	170 ve daha çok	14



FREKANS POLİGONLARI

- Bu grafiklerde sınıf aralığı yerine sınıf orta noktasının sınıf frekansına göre dağılımı çizgisel olarak grafiklenir.
- Frekans poligonlarının histogramlara göre bir avantajı 2 veya daha fazla frekans dağılım grafiğinin kolaylıkla birbirleri ile karşılaştırılmasına imkan tanımasıdır (Frekans dağılımlarının sınıf sayısı ve aralığı aynı olmak koşulu ile)



II. Merkezi Eğilim Ölçüleri

Doç. Dr. İrfan Yolcubal
Kocaeli Üniv. Jeoloji Müh. Böl.

Analitik Ortalamalar

- Bir örneklemede tüm veri değerlerini dikkate alan merkezi eğilim ölçüleridir.
  - Aritmetik ortalama
  - Ağırlıklı ortalama
  - Geometrik ortalama
  - Harmonik ortalama

Aritmetik Ortalama (Aritmetic mean)

Gruplanmamış yani ham veriler için aritmetik ortalama, tüm veri değerlerinin toplamının toplam veri sayısına oranıdır.

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{N} = \frac{X_1 + X_2 + X_3 + ... X_n}{N}$$

$\overline{X}$: Aritmetik ortalama
N: Toplam veri sayısı
X: veri değeri

Tekrarlanan gözlemlerin veri değerlerinin hesabı

- Bir örnekte gözlem değerleri bir çok kez tekrarlanabilir. Örneğin, 20 kişilik bir sınıfta istatistik dersinden geçen öğrencilerin notu şu şekilde sıralanmaktadır: 5, 5, 5, 5, 5, 6, 6, 6, 7, 7, 8, 8, 8, 9, 9, 9, 10, 10, 10, 10 Bu sınıfın istatistik dersinin aritmetik ortalaması nedir?

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} f_i X_i$$

$$n = \sum_{i=1}^{k} f_i$$

f_i: bir örnekteki X_i`nin frekansı
k: örnekteki gözlem sayısı
X_i: i.gözlem değeri
n=toplam veri sayısı

$$\overline{X} = \frac{(5*5)+(3*6)+(2*7)+(3*8)+(3*9)+(4*10)}{20} = 7.4$$

Aritmetik Ortalamanın Özellikleri

- Bir serideki her bir veri değerinin aritmetik ortalamadan olan sapmalarının toplamı daima sıfırdır.

$$\sum(X-\overline{X})=0$$

Örnek: 3, 8, ve 4 değerlerin aritmetik ortalaması 5`dir.

$$\sum(X-\overline{X})=(3-5)+(8-5)+(4-5)=-2+3-1=0$$

•Aritmetik ortalamanın hesaplanışında veri setindeki tüm veri değerleri kullanılır.

•Bir veri setinin yalnızca bir aritmetik ortalaması vardır.

Aritmetik Ortalamanın Dezavantajı

- Aritmetik ortalamanın dezavantajı, veri setindeki aşırı değerlerden kolay etkilenmesidir. Bir veri setindeki verilerden bir kaçı çok yüksek yada düşük değerler içeriyor ise, aritmetik ortalama, veri setinin merkezi eğilim ölçümünü temsil etmek için uygun olmayabilir.
  - Örnek: 5 öğrencinin bir sınavda almış olduğu notlar 70, 70, 70, 70, ve 100`dir. Aritmetik ortalama 76 olacaktır. Bu aritmetik ortalama veri setinin iyi bir şekilde temsil etmemektedir.
- Açık sınıf aralıklı frekans dağılım tablolarında aritmetik ortalama uygun değildir.

Ağırlıklı Ortalama(Weighted Mean)

- Aritmetik ortalamada, her bir veri değerinin öneminin eşit olduğu varsayılmaktadır. Fakat bazı değerlerin önemi diğerlerinden farklı olabilir. Bu durumlarda ağırlıklı ortalama kullanılır.

$$\overline{X}_W = \frac{X_1W_1 + X_2W_2 + X_3W_3 + \ldots\ldots + X_nW_n}{W_1 + W_2 + W_3 + \ldots\ldots + W_n}$$

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

W: Her bir veri değerinin ağırlığını yani önemini ifade etmektir.

$\overline{X}_W$: Ağırlıklı ortalama

Ağırlıklı Ortalama`ya Örnek

- Bir öğrenci matematik dersinden 6, edebiyat dersinden 7, müzik dersinden 10 ve İngilizce dersinden 9 almıştır. Bu öğrencinin ders ortalamasını hesaplayalım.
- Ders kredileri: Matematik:3, edebiyat: 2, müzik:1, İngilizce: 1

$$\overline{X}_W = \frac{(6*3)+(7*2)+(9*1)+(10*1)}{3+2+1+1} = \frac{51}{7} = 7.3$$

$$\overline{X} = \frac{6+7+9+10}{4} = 8$$

Ağırlıklı Ortalama: Örnek 2

- Sarar normal perakende satış fiyatı üzerinden ($400) 95 adet Kığılı marka takım elbiseyi satmıştır. Bahar indiriminde aynı takım elbiseyi 200 dolara indirerek 126 adet, en son indirimde de 100 dolara 79 adet takım elbise satmıştır. Sarar takım elbisesinin ağırlıklı ortalama fiyatı nedir? Sarar her bir takım elbiseye birim fiyatı olarak $200 ödemiştir. Sararın bu satıştaki toplam kazancı ne kadardır?

Örnek 2

$$\overline{X}_W = \frac{(95 * 400) + (126 * 200) + (79 * 100)}{95 + 126 + 79}$$

$$\overline{X}_W = \frac{71100}{300} = \$\,237$$

$$\textit{Birim Kazanç} : 237 - 200 = \$\,37$$

$$\textit{Toplam Kazanç} : 37 * 300 = \$\,12100$$

Geometrik Ortalama (Geometric Mean)

- Geometrik ortalama iktisat ve işletme alanlarında yaygın olarak kullanılan bir ortalama türüdür. Geometrik ortalama özellikle

  1) değişim oranlarının (yüzde, oran, vb.) ortalamasının hesaplanmasında 2) bir zaman aralığı içerisindeki bir üretimin yada satışın artış miktarının ortalamasının belirlenmesinde yaygın olarak kullanılmaktadır.

$$G.O. = \sqrt[n]{(X_1)(X_2)(X_3).....(X_n)}$$

Not: Eğer veri değerlerinden bir 0 yada negatif değerlikli ise Geometrik ortalama hesaplanamaz.

$$\log \; G.O. = \frac{1}{n}\sum \log \; X_i$$

Veri sayısı çok olduğu durumlarda hesapları kolaylaştırmak amacıyla logaritmalardan yararlanılmaktadır.

Geometrik Ortalama: Örnek 1

Bir inşaat şirketinin dört projedeki ortalama kâr yüzdeleri 3, 2, 4, Ve 6`dır. Bu şirketin ortalama kârı nedir?

$$G.O. = \sqrt[4]{3*2*4*6} = \sqrt[4]{144} = \%3.46$$

$$\overline{X} = \frac{3+2+4+6}{4} = \%3.75$$

Geometrik ortalama daha tutucu bir kar değeri vermektedir. Çünkü aşırı değerlerden aritmetik ortalamaya göre o kadar fazla etkilenmemektedir. Bu nedenle geometrik ortalama ya aritmetik ortalamaya eşit olacaktır yada küçük olacaktır.

Geometrik Ortalama: Örnek 2

- Türkiye'nin nüfusu 1990 yılında 50.7 milyondan 1995 yılında 56.5 milyona yükselmiştir. Bu 5 yıl içinde nüfusun ortalama artış hızı ne olmuştur?

$$G.O. = \sqrt[n-1]{\frac{bir\,periyotun\ sonundaki\ \text{deger}}{\text{bir periyotun baslangicindeki deger}}} - 1$$

n : periyot araligi

$$G.O. = \sqrt[5-1]{\frac{56.5}{50.7}} - 1 = \sqrt[4]{1.11} - 1 = \%2.47$$

Harmonik Ortalama(Harmonic mean)

- Bazı özel durumlarda başvurulan bir ortalama olup hız, fiyat, verimlilik gibi oransal olarak belirtilebilen bazı değişken değerlerin ortalamalarının hesaplanışında kullanılır.

$$H = \frac{N}{\frac{1}{X_1} + \frac{1}{X_2} + \frac{1}{X_3} + ... + \frac{1}{X_n}} = \frac{N}{\sum \frac{1}{X_i}}$$

•Değişkenlerden birinin sabit, diğerinin ise değişken olduğu durumlarda başvurulan bir ortalamadır.

•Veri değerlerinde sıfır bulunması yada veri değerlerinin farklı işaret taşımaları durumunda harmonik ortalama kullanılmaz.

Harmonik Ortalama: Örnek

•İki kasaba arasındaki mesafe gidişte saatte 75 km. hızla, dönüşte ise 50 km hızla kat edilmektedir. Bu durumda ortalama hız nedir?

İki kasaba arasındaki mesafe 150 km varsayılır ise gidiş için gerekli süre 150/75: 2 saat, dönüş için ise 150/50: 3 saat

Burada mesafe unsuru sabit fakat zaman unsuru ise sabit olmadığından harmonik ortalama kullanılmıştır.

$$H = \frac{2}{\frac{1}{75} + \frac{1}{50}} = 60\ km$$

$$\cancel{\overline{X} = \frac{75 + 50}{2} = 62.5km}$$

ANALİTİK OLMAYAN MERKEZİ EĞİLİM ÖLÇÜLERİ

- Bir örnekteki bütün veri değerlerini dikkate almayan merkezi eğilim ölçüleridir.
  - Medyan (Median)
  - Mod (Mode)

Medyan (Ortanca)

- Bazı durumlarda örneğin bir yada iki tane çok yüksek yada düşük değerler içerebileceğinden bahsetmiştik. Bu gibi durumlarda aritmetik ortalama örneğin merkezi eğilimini yansıtmaz. Böyle problemlerde medyan değeri kullanılarak örneğin merkezi eğilimi ölçülebilir.
- Veri değerleri büyükten küçüğe yada küçükten büyüğe sıralandıktan sonra, tam ortadaki yani veri dizisini 2 eşit frekansa ayıran değerdir.
- Düzenlenmemiş verilerde medyan`ın yerini kolaylıkla tespit etmek için aşağıdaki formülden yararlanılabilir.

$$\text{medyan degerinin yeri} = \frac{n+1}{2}$$

$$n = \text{toplam veri sayisi}$$

Medyan: Örnek1

- Bir klinikte pansuman için ödenen miktarlar aşağıdaki gibi sıralanmaktadır: 65, 29, 30, 25, 32, 35 TL. Medyan fiyat nedir?

25
29
30
32 ← Medyan : (30+32)/2= 31 TL.
35
65

Medyan: Örnek 2

- Yuvacık Kalıcı konutlarındaki kira fiyatları aşağıdaki gibi sıralanmaktadır: 120,100, 110, 115, 125, 105, 70 TL. Ortalama kira fiyatı nedir.

Medyan: Örnek 2

70
100
105
110 ← Medyan
115
120
125

Medyanın Özellikleri

- Her bir veri setinin tek bir medyanı vardır.
- Veri setindeki aşırı değerlerden etkilenmediği için verilerin merkezi eğiliminin belirlenmesinde aritmetik ortalamaya nazaran daha doğru bir bilgi sunar. Aritmetik ortalamanın aksine açık sınıf aralıklı frekans dağılımlarının merkezi eğiliminin ölçümünde kullanılabilirler.

Mod

- Bir veri setindeki bütün değerleri dikkate almayan (hassas olmayan) bir başka merkezi eğilim ölçümüdür.
- Mod, bir data setinde en sık olarak gözlenen veri değeridir.

Mod: Örnekler

- 4, 6. 5, 8, 7, 10, 9, 11 Mod ?
- 4, 6, 5, 4, 7, 5.5, 4, 6.5, 7, 8, 4, 6, 4, 5, 4, 4 Mod ?
- 4, 6, 4, 5, 6, 5, 6, 6, 5, 6, 5, 5 Mod ?

Gruplanmış Verilerde Aritmetik Ortalama, Medyan, Mod

- Bazı durumlarda veri değerleri gruplandırılıp, frekans dağılımları oluşturulmuş olabilir ve ham veriler mevcut bulunmayabilir. Bu gibi durumlarda aritmetik ortalama, medyan ve mod frekans dağılım tablolarından hesaplanabilir.
- Bu değerler gerçek ham verilerden hesaplanan değerlerden farklı olabilir.

Aritmetik Ortalamanın Frekans Dağılımından Hesaplanışı

$$\overline{X} = \frac{\sum fX}{N}$$

X: her bir sınıfın orta noktası
f: her bir sınıf frekansı
N: Toplam veri sayısı yada frekansların toplam değeri

Örnek

Net geliri (milyon $)	İthalatçı sayısı
2-5	1
5-8	4
8-11	10
11-14	3
14-17	2

Net gelirin aritmetik ortalamasını hesaplayınız?

Net geliri (milyon $)	İthalatçı sayısı (f)	Sınıf orta noktası (X)	fX
2-5	1	3.5	3.5
5-8	4	6.5	26
8-11	10	9.5	95
11-14	3	12.5	37.5
14-17	2	15.5	31
Toplam	20		193

$$\overline{X} = \frac{\sum fX}{N} = \frac{193}{20} = 9.65$$

Frekans Dağılımından Medyanın Hesaplanışı

$$Medyan = L + \frac{\frac{N}{2} - CF}{f}(i)$$

L: Medyan sınıfın alt sınırı
N: Toplam frekans değeri
CF: Medyan sınıfından önceki sınıfların frekanslarının toplamı
f: medyan sınıfının frekansı
i: medyan sınıfının aralığı

Net geliri (milyon $)	İthalatçı sayısı (f)	Sınıf orta noktası (X)	CF
2-5	1	3.5	1
5-8	4	6.5	5
8-11	10	9.5	15
11-14	3	12.5	18
14-17	2	15.5	20
Toplam	20		

$$medyan = 8000000 + \frac{\frac{20}{2} - 5}{10}(3000000)$$

$$medyan = 9500000$$

Frekans Dağılımından Modun Hesaplanışı

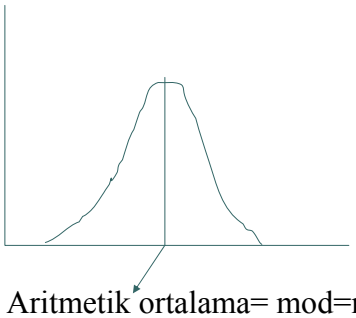
- Frekans dağılımı şeklinde gruplanmış veriler için mod, frekans sayısı en fazla olan sınıfın orta noktası değeridir.
- Eğer frekans dağılımında 2 sınıf maksimum frekansa sahip ise bu tür dağılımlara bimodal dağılımlar denilmektedir.

Bir ürünün satış fiyatın gruplandırılmış şekli.

Net satış ($ dolar)	Toplam yüzdesi
1-4	13
4-7	14
7-10	40
10-13	23
13 ve daha fazlası	10

Net satışın medyan ve modu hesaplayınız? Bu tür tablolara ne ad verilmektedir.

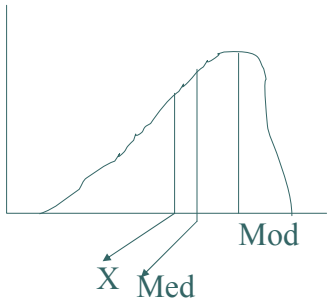
Bir Frekans Dağılımı Grafiğinde Aritmetik ortalama, medyan, mod arasındaki ilişki



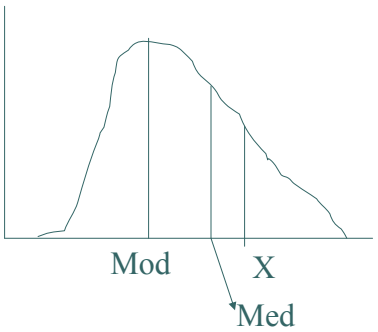
Aritmetik ortalama= mod=medyan

Simetrik frekans dağılım eğrileri

Bir Frekans Dağılımı Grafiğinde Aritmetik ortalama, medyan, mod arasındaki ilişki



Asimetrisi negatif olan
Frekans dağılımları(sola çarpık)



Asimetrisi pozitif olan frekans
Dağılımları(sağa çarpık)

Tek modlu ve asimetrileri çok fazla olmayan veriler için
Aşağıdaki ilişki kullanılabilir.
Aritmetik ortalama-Mod= 3(Aritmetik Ort-Medyan)

III. DAĞILMA YADA DEĞİŞKENLİK ÖLÇÜLERİ (MEASURE OF DISPERSION)

Doç. Dr. İrfan Yolcubal
Kocaeli Üniv.
Jeoloji Müh. Böl.

Dağılma (değişkenlik) ölçülerinin analizinin nedeni

- Bir veri setinin merkezi eğilim ölçüsünün değerlendirilmesi
- 2 veya daha fazla veri setinin dağılımının karşılaştırılması

Dağılma yada değişkenlik ölçüleri

- Değişkenlik aralığı (Range)
- Ortalama sapma (mean deviation)
- Varyans (Variance)
- Standard sapma (Standard deviation)

Gruplanmamış verilerde dağılımın ölçümü

- Değişkenlik aralığı (Range): Bir veri serisindeki en yüksek değer ile en düşük değer arasındaki farka eşittir.

R= Maksimum değer- Minimum değer

Değişkenlik aralığı (Range)

Seri 1	Seri 2
2	5
3	5
6	5
7	6
8	7
10	8
$\overline{X} = 6$	$\overline{X} = 6$
R=10-2=8	R=8-5=3

Gruplanmış verilerde değişkenlik aralığının(Range) belirlenmesi

Saatlik Ücret(lira)	Frekans
5-10	10
10-15	21
15-20	9
20-25	5

R= 25-5=20 TL

Ortalama yada Mutlak Ortalama Sapma (Mean Deviation or Mean Absolute Deviation)

- Bir popülasyondaki tüm veri değerlerinin popülasyonun aritmetik ortalamasından olan mutlak sapmalarının aritmetik ortalamasıdır.

$$O.S. = \frac{\Sigma\left|X - \overline{X}\right|}{N}$$

$\left|X - \overline{X}\right| =$ mutlak sapma

Ortalama Sapma (Gruplanmamış Verilerde)

- 15, 16, 18, 21, 25 değerlerinden meydana gelmiş serinin ortalama sapmasını hesaplayalım.

Veriler	$\left\|X - \overline{X}\right\|$	$\left\|X - \overline{X}\right\|$
15	| 15-19 |	4
16	| 16-19 |	3
18	| 18-19 |	1
21	| 21-19 |	2
25	| 25-19 |	6
	TOPLAM	16

$$O.S. = \frac{16}{5} = 3.2$$

Ortalama sapma (Gruplanmış Verilerde)

Sınıf	frekans (f)	Sınıf orta noktası (X)	fx	$\lvert X - \overline{X} \rvert$	$\lvert X - \overline{X} \rvert$	$f \lvert X - \overline{X} \rvert$
10-20	7	15	105	|15-33.2|	18.2	127.4
20-30	14	25	350	|25-33.2|	8.2	114.8
30-40	16	35	560	|35-33.2|	1.8	28.8
40-50	9	45	405	|45-33.2|	11.8	106.2
50-60	5	55	275	|55-33.2|	21.8	109
Toplam	***51***		***1695***			***486.2***

$$\overline{X} = \frac{\sum fX}{N} = \frac{1695}{51} = 33.2$$

$$O.S. = \frac{\sum f \lvert X - \overline{X} \rvert}{N} = \frac{486.2}{51} = 9.53$$

Varyans (Variance) ve Standart sapma (Standard deviation)

- Varyans, bir veri setindeki tüm veri değerlerinin, ortalamadan olan sapmaların karesinin aritmetik ortalamasıdır.

$$\sigma^2 = \frac{\Sigma \left(X - \overline{X}\right)^2}{N}$$

Popülasyona ait

$$\sigma^2 = \frac{\Sigma \left(X - \overline{X}\right)^2}{N-1}$$

Örneklemeye ait

Standart Sapma

- Bir veri setinin varyansının kareköküne eşittir.

$$\sigma = \sqrt{\frac{\Sigma\left(X - \overline{X}\right)^2}{N}}$$

Popülasyona ait

$$\sigma = \sqrt{\frac{\Sigma\left(X - \overline{X}\right)^2}{N-1}}$$

Örneklemeye ait

Standart sapma ve varyansın özellikleri

- Bir serideki değerlere bir sabitin (a'nın) eklenmesi veya bu değerlerden bir sabitin çıkarılması ile serinin varyans ve standart sapması değişmez.

$$\sigma^2(X - a) = \sigma^2(X)$$
$$\sigma^2(X + a) = \sigma^2(X)$$

- Bir serideki değerlerin her birinin belirli bir sabit (c) ile çarpılması sonucu meydana çıkan serinin varyansı, orijinal serinin varyansının sabitin karesi ile çarpımına eşittir. Ancak standart sapması, orijinal serinin standart sapmasının sabitin kendisi ile çarpımına eşit olmaktadır.

$$\sigma^2(cX) = c^2\sigma^2(X)$$
$$\sigma(cX) = c\sigma(X)$$

Standart Sapma ve Varyans (Gruplanmamış verilerde)

- 22, 25, 28, 30, ve 35 değerlerinden oluşan popülasyonun varyans ve standart sapmasını hesaplayalım.

X	$(X-\overline{X})$	$(X-\overline{X})^2$
22	-6	36
25	-3	9
28	0	0
30	2	4
35	7	49
$\sum X = 140$	$\sum(X-\overline{X}) = 0$	$\sum(X-\overline{X})^2 = 98$

$$\overline{X} = \frac{140}{5} = 28$$

$$\sigma^2 = \frac{98}{5} = 19.6$$

$$\sigma = \sqrt{19.6} = 4.43$$

Standard Sapma ve Varyans Hesabı (Gruplanmış verilerde)- Popülasyon örneği

sınıflar	frekans (f)	Sınıf orta noktası (X)	fX	$(X-\overline{X})$	$(X-\overline{X})^2$	$f(X-\overline{X})^2$
0-200	8	100	800	-270	72900	583200
200-400	11	300	3300	-70	4900	53900
400-600	7	500	3500	130	16900	118300
600-800	6	700	4200	330	108900	653400
Toplam	***32***		***11800***			***1408800***

$$\overline{X} = \frac{11800}{32} = 370$$

$$\sigma^2 = \frac{1408800}{32} = 44025$$

$$\sigma = \sqrt{44025} = 209.82$$

Gruplanmış verilerin varyansı ve standart sapması

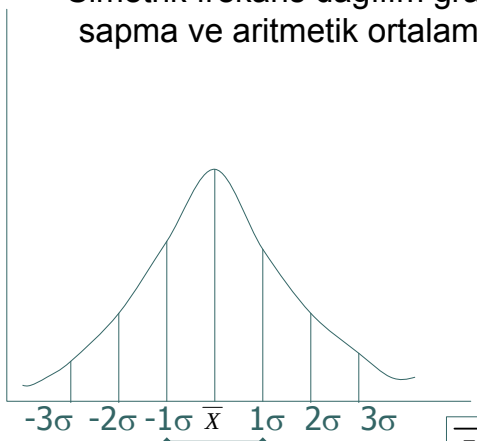
- Sınıflandırılmış verilerde, sınıf orta noktası her zaman bu kısmın ağırlıklı orta noktası olmayacağından, ham verilere göre gruplandırılmış değerlerde daha yüksek varyans değeri bulunur. Bunu önlemek için Sheppard düzeltmesi yapılır.

- $Sh=c^2/12$
  - c=sınıf aralığı
  - Sh=sheppard düzeltmesi
- Düzeltilmiş varyans = σ^2-Sh

Varyans

- Tek bir veri seti için varyans değerinin yorumlanması zordur. Değişkenlik aralığı ve ortalama sapmada olduğu gibi varyans, 2 veya daha fazla data setindeki verilerin değişkinlik derecelerin karşılaştırılmasında kullanılır.
-

Simetrik frekans dağılım grafiklerinde standart sapma ve aritmetik ortalama arasındaki ilişki



%68 → $\overline{X} \pm 1\sigma$

%95 → $\overline{X} \pm 2\sigma$

%99.7 → $\overline{X} \pm 3\sigma$

Nisbi Dağılma(Relative Dispersion)

- Değişim katsayısı (Coefficient of variation): % olarak ifade edilir. Bir data setinin standart sapmasının aritmetik ortalamasına oranıdır.

$$\text{Degisim Katsayisi}(\%) = \frac{\sigma}{\overline{X}} * 100$$

  - Birimleri farklı olan farklı 2 data setinin kıyaslanmasında (cm, TL)
  - Birimleri aynı fakat ortalamaları birbirinden çok farklı olan data setlerin kıyaslanmasında değişim katsayısı kullanılır.

$\overline{X} = 500000, \sigma = 50000$

$\overline{X} = 12000, \sigma = 2000$

$$\text{Degisim katsayisi} = \frac{50000}{500000} * 100 = 10\%$$

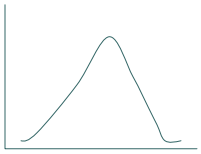
$$\text{Degisim katsayisi} = \frac{2000}{12000} * 100 = 16.7\%$$

Gruplanmamış verilerde Çarpıklık

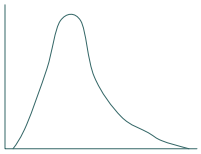
- Frekans dağılım grafiklerinin çarpıklık derecesini tarif etmektedir.

$$Çk = \frac{\sum \left(x_i - \bar{x}\right)^3}{n\sigma^3}$$

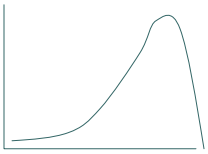
- Çarpılık genellikle –3 ile +3 değerleri arasında değişmektedir.



Simetrik çarpıklık=0



Pozitif çarpıklık



Negatif çarpıklık

Gruplanmış verilerde Çarpıklık

$$Çk = \frac{\sum f_i \left(x_i - \bar{x}\right)^3}{\sum f_i \sigma^3}$$

$$\boxed{\text{Carpiklik} = \frac{3(\bar{X} - \text{medyan})}{\sigma}}$$ Şeklinde de hesaplanabilir.

Gruplanmamış verilerde Kurtosis (Basıklık-Sivrilik)

- Frekans dağılım eğrisinin yüksekliğini belirten bir parametre.

$$Kur = \frac{\sum \left(x_i - \bar{x}\right)^4}{n\sigma^4}$$ Ham verilerde

$$Kur = \frac{\sum f_i \left(x_i - \bar{x}\right)^4}{\sum f_i \sigma^4}$$ Gruplanmış verilerde

- Kur=3 (mesokurtik, yada normal dağılım)
- Kur>3 (Leptokurtik, sivri-orta kısımlar uç kısımlara göre daha İyi boylanmış yada sivrileşmiş can eğrisi
- Kur<3 (Platykurtik,basık uç kısımlar daha iyi boylanmış yada basık
- Çan eğrisi

IV. OLASILIK YADA İHTİMAL TEORİSİ

Dr. Irfan Yolcubal
Kocaeli Üniversitesi

Olasılık Kavramı

- Olasılık (Probability): Belirli bir olayın olma ihtimalinin yada şansının ölçümü.
  - $0 \leq P(A) \leq 1$. Oran olarakta ifade edilebilir (Örneğin, 7/10; 20/100; 1/2)
  - 0 olasılık olayın kesinlikle olmayacağını; 1 olasılık ise olayın kesinlikle meydana geleceğini ifade etmektedir.

Terminaloji: Deney (Experiment), Sonuç (Outcome), ve Olay (Event)

Deney	Sonuç	Olay
Zar atma	1, 2, 3, 4, 5, 6	▪ Çift bir sayının gelmesi ▪ 4`den büyük sayının gelmesi, vb.

Deney: Bir aktivitenin gözlemlenmesi yada ölçüm alma şekli
Sonuç: Bir deneyin belli bir sonuçu
Olay: Bir deneyin bir yada daha fazla sonuçlarının toplamı

Olasılık Sınıflandırılması

- Olasılık 2 şekilde sınıflandırılabilir.
- Objektif Olasılık (Tekrarlanabilen rastgele bir deneye bağlı). Örnek: Rus ruletinin döndürülmesi, zar atılması
  - Klasik Olasılık
  - Nisbi frekans olarak 2 kısma ayrılabilir.
- Subjektif Olasılık (Tekrar edilemeyen bir deneye bağlı). Örnek: Geçmis meteoroloji verilerine dayanarak yarın yağmur yağma ihtimalinin tahmini subjektif olasılık değerleri sunan bir deneydir.

Klasik Olasılık Tanımı

- Eğer bir olay E, mümkün olan ve ortaya çıkma şansı eşit olan bütün hallerden (n) sadece (a) kadar halde ortaya çıkıyor ise, o olayın ihtimali

$$P(E) = \frac{a}{n} = \frac{\text{elverisli hal sayisi}}{\text{toplam mumkun haller sayisi}}$$

- Örnek: Bir zar atıldığında 2 gelme olasılığı, P=1/6 = 0.167

Klasik Olasılık Kavramı

- Olayın çıkışını elverişli hal(a), çıkmayışınıda elverişsiz hal (b) olarak ifade edersek,
  - n= a+b
- Elverişsiz bir halin ortaya çıkma olasılığı [P(~E), yada $P(\overline{E})$]

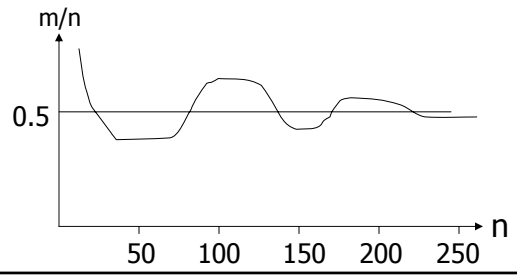
$$P(\overline{E}) = \frac{b}{n} = \frac{n-a}{n} = 1 - \frac{a}{n} = 1 - P(E)$$

Örnek: Bir zar atıldığında 2 gelmeme olasılığı nedir?
P(~E)=1-P(E)= 1-1/6 =0.833

Olasılığın nisbi frekans olarak belirlenmesi

- Bir olayın meydana gelme olasılığı= geçmişte benzer olayın tekrarlanma sayısının toplam gözlem sayısına oranıdır.

Örnek. 250 kez para atıldığında yazı gelme olaylarının nisbi frekansını hesaplarsanız



$$P(\text{Yazı}) = m/n \quad n \longrightarrow \infty$$

m: yazı sayısı
n: toplam atış

OLASILIK HESAPLAMARINDA TEMEL KURALLAR

- Olasıkların toplanması: Bileşik bir olayın ortaya çıkma olasılığı, o olayı meydana getiren olayların olasılıkların toplamına eşittir.
- Örneğin A veya B olayının meydana gelme olasılığı

$$P(A \text{ veya } B) = P(A) + P(B)$$

- Bu kuralın uygulanabilmesi için bir olay meydana gelirken aynı anda diğerlerinin meydana gelmemesi gerekmemektedir. Yani bileşik olayların birbirlerini engelleyen türden olması gerekmektedir [P(AB)=0].
- Bir kere zar atıldığında 2 yada 6 gelme olasılığı nedir? P(2 yada 6) = 1/6 +1/6 =0.33

Olasılıkların toplanması

- Eğer bir deneyde A, B, ve C olayları birbirlerini engellemeyen türden olaylar ise A ve B olaylarının ortaya çıkış olasılığı
- P(A veya B)= P(A) +P(B)-P(AB)
  - P(AB): Birleşik olasılık (Joint probabilit):hem A hemde B olayının aynı anda meydana gelme olasılığı
- A veya B veya C'nin ortaya çıkış olasılığı;

P (A veya B veya C)= P(A) + P(B)- P(AB)-P(AC)-P(BC)+P(ABC)

Örnek 1

- Bir deste iskambil kağıtı arasından 1 vale veya 1 maça çekme olasılığı nedir?

Vale çekme olayı: A; maça çekme olayı: B

P(A veya B)=P(A)+P(B)-P(AB)

= 4/52 + 13/52 –1/52 = 4/13

Örnek 2

- Ülkemizi ziyaret eden 200 turist üzerinde yapılan ankette sadece Ayasofyayı ziyaret eden turist sayısı 120, Efesi ziyaret eden turist sayısı ise 100'dür. 60' ise her ikisinide ziyaret etmiştir. Bir turistin Ayasofya veya efesi ziyaret etme olasılığı nedir?

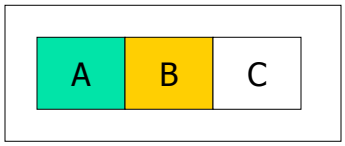
Ayasofyayı ziyaret etme olayı: A; efesi ziyaret etme olayı:B

P(A veya B)=P(A) + P(B)- P(AB)

- =120/200 +100/200 –60/200= 0.8

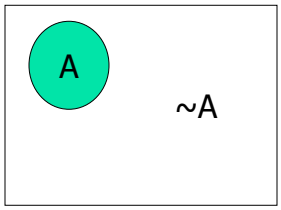
Olasılıklarınının Toplanmasının Küme Teorisi ile açıklanması

- J. Venn(1834-1888) bir deneyin sonuçlarını grafiksel olarak ifade edebilmek için diyagramlar geliştirdi. Bu diyagramlara Venn diyagramları denir.



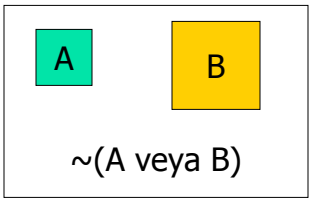
A, B, C olayları birbirlerini engeleyen olaylar

P(A)+P(B)+P(C) =1

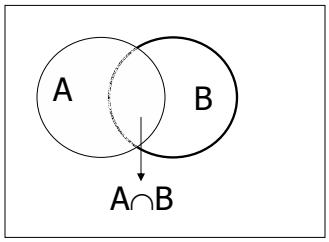


P(A)+P(~A)=1

Olasılıklarınının Toplanmasının Küme Teorisi ile açıklanması



P(A veya B)= P(A)+P(B)



P(A veya B)=P(A)+P(B)-P(AB)

Olasılıkların çarpımı kuralı

- Şartlı olasılık(Conditional Probabilility): Şayet A olayı B olayından sonra ortaya çıkıyorsa, A olayının olasılığına şartlı olasılık denir ve P(A|B) şeklinde ifade edilir.
- P(A|B): Eğer B olayı meydana geldiyse A olayının olasılığını ifade etmekte.

$$P(A|B) = \frac{P(AB)}{P(B)}$$

Ave B olayları birbine bağlı ise yani B olayının ortaya çıkışı A olayının olasılığını etkiliyorsa
hem A hemde B olayının aynı anda meydana gelme olasılığı

$$P(AB) = P(A).P(B|A)$$
$$P(AB) = P(B).P(A|B)$$

Örnek

- Bir kutuda 10 adet film var. Bunlardan 3'ünün bozuk olduğu biliniyor. Eğer sırasıyla birer adet toplam 2 film çekersek, seçilmiş filimlerin her ikisininde bozuk olma olasılığı nedir?
- P(A): birinci çekimde filmin bozuk çıkma olasılığı
- P(B): ikinci çekimde filmin bozuk çıkma olasılığı
- P(A ve B)= P(A).P(B|A)

  =3/10 . 2/9 =0.0667

Olasılıkların çarpımı kuralı

- Eğer A ve B olayları birbirinden bağımsız ise yani birinin ortaya çıkışı diğerinin olasılığını etkilemiyor ise

Hem A hemde B'nin aynı anda ortaya çıkma olasılığı

$$P(AB)=P(A)P(B)$$

Ödev

- Eğer A, B, ve C olayları birbirine bağlı ise A,B,ve C'nin aynı anda ortaya çıkma olasılığını nasıl hesaplarız?

Bayes Kuralı

- Şartlı olasılıkların hesaplanmasında kullanılan bir tekniktir. Kuralın amaçı bir olayın ortaya çıkmasında birden fazla bağımsız nedenin etkili olması halinde bu nedenlerden herhangi birinin, o olayı yaratmış olması ihtimalini hesaplayabilmektir.

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1)+\ldots+P(A_n)P(B|A_n)}$$

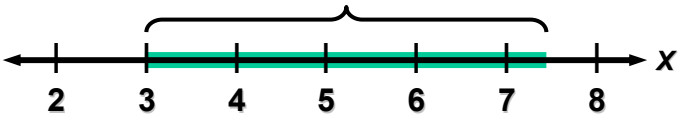
SÜREKSİZ(DISCRETE) OLASILIK DAĞILIMLARI

Yrd. Doç.Dr. İrfan Yolcubal
Kocaeli Üni. Jeoloji Müh.

- Random Değişken: Nümerik olarak ifade edilen bir deneyin sonuçları
- Süreksiz(Discrete) Random Değişken:
  - Random değişken belirli bir aralıkta sadece kesin değerler alabilir.
  - Random değişkenin değeri sayarak elde edilir.
  - Örnek:
    - Bir parayı 5 kez atalım ve tura gelme sayısını hesaplayılım (0, 1, 2, 3, 4, yada 5)
    - Bir haftada yapılan satış sayısı
    - 5 dakika içerisinde Türk petrole gelen araç sayısı
    - 200 kişi içerisinden uçak rezervasyonu yaptırıp, sonra vazgeçen kişi sayısı

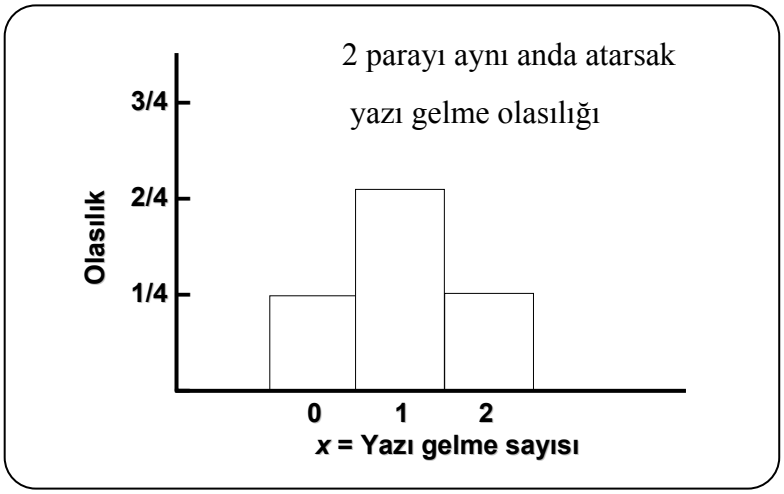
Sürekli (Continuous) Random Değişken:

Random değişken belirli bir aralıkta her değeri alabilir.



Süreksiz Random Değişkenlerin Olasılık Dağılımları

- Olasılık dağılımı, bir deneyin olası sonuçlarını ve her bir sonuça karşılık gelen olasılıkları sunan dağılımlardır.



Süreksiz(Discrete) Olasılık Dağılımları

- Tüm olası [x_i , p(x_i)] çiftleri
- x= random değişkeninin değeri
- P(x)= random değişkenin meydana gelme olasılığı

$$0 \leq p(x_i) \leq 1$$

- $$\Sigma\, P(x_i) = 1$$
- Deneyin sonuçlarına ait olasılıkların toplamı 1'e eşittir.
- Deneyin sonuçları birbirlerini engellemeli

Süreksiz Olasılık Dağılımlarının ortalaması ve varyansı

$\mu = E(X) = \sum x_i\, p(x_i)$

$\sigma^2 = \Sigma\, (x_i - \mu)^2 p(x_i)$

Örnek: 2 parayı aynı anda atalım ve yazı gelme sayısını belirleyelim. Ortalama değeri ve varyansı hespalay

$\mu = 0 \times .25 + 1 \times .50 + 2 \times .25 = 1$

$\sigma^2 = (0 - 1)^2(.25) + (1 - 1)^2(.50) + (2 - 1)^2(.25) = .50$

Süreksiz Olasılık Dağılım Modelleri

Süreksiz Olasılık Dağılımları

- **Binom**
- **Hipergeometrik**
- **Poisson**

BİNOM OLASILIK DAĞILIMLARI

- Binom olasılık dağılımları n adet deneyde kesin olarak x adet elverişli halin meydana gelme olasılığını tarif eden dağıılımlardır.
- Binom dağılımlarında n adet tekrarlanan benzer deney vardır.
- Deneyin her bir sonuçu sadece 2 olaydan biri olabilir: Bunlar genelde elverişli(success) ve elverişsiz (failure) olaylar olarak sınıflandırılır.
- Deneyin her bir sonuçun olasılığı bir deneyden diğerine değişmez.
- Tekrarlanan deneylerde, bir deneyin sonuçu diğerlerinkini etkilemez yani bağımsızdır.
- Örnek: Bir paranın 15 kez yazı/tura için atılması, bir nalburdan 10 adet lamba alınması vb..

BİNOM OLASILIK DAĞILIMLARI

$$P(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

n: deney sayısı
x: elverişli halin sayısı
p: her bir deneydeki elverişli halin olasılığı
q: elverişsiz halin olasılığı, q=1-p

Bir para 2 kere atıldığında tura gelme olasılığı

X	P(X)
0	1/4 = .25
1	2/4 = .50
2	1/4 = .25

Her x değeri için

$\mu = E(x) = np$

$\sigma^2 = E[(x - \mu)^2] = np\,(1 - p)$

Binom Olasılık Dağılımları ortalama değer ve varyans

Ortalama

$$\mu = E(X) = np$$

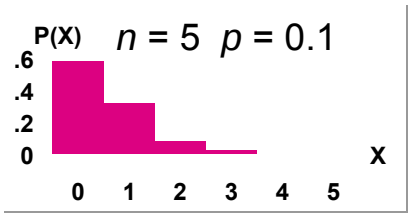
e.g. $\mu = 5\,(.1) = .5$

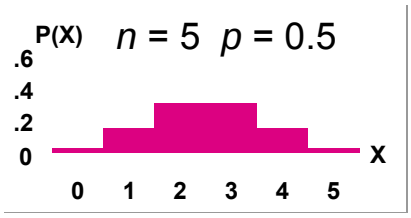
Standard Sapma

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

e.g. $\sigma = \sqrt{5(.5)(1 - .5)}$

$= 1.118$



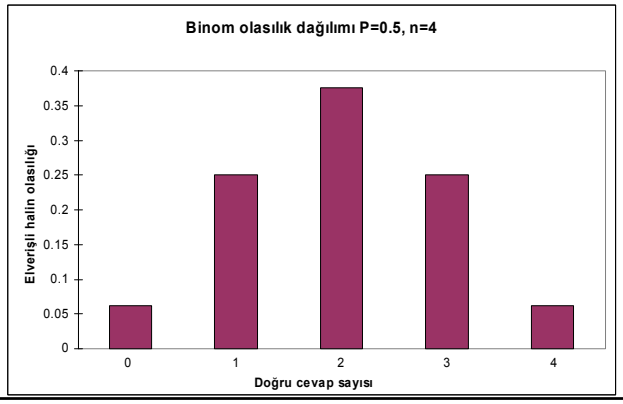


Örnek: Binom Dağılımları

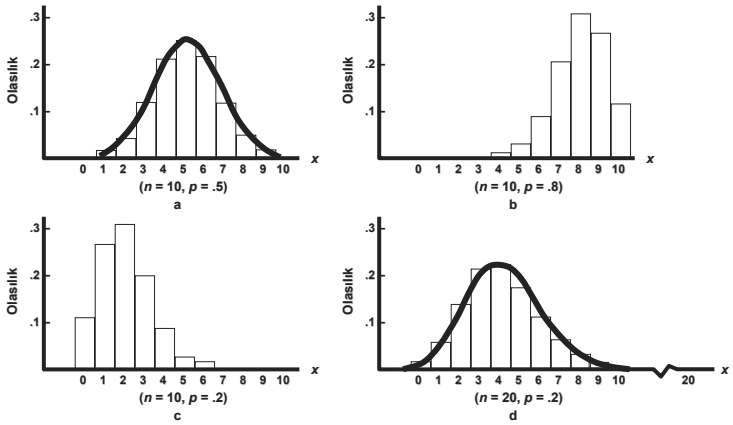
- 4 D/Y sorusundan 0, 1 doğru çevap bulma olasılığı nedir? Not: Ögrenciler konu hakkında önceden bir bilgiye sahip değildirler.

$$P(0) = \frac{4!}{0!(4-0)!}(0.5)^0(1-0.5)^{4-0} = 0.0625$$

$$P(1) = \frac{4!}{1!(4-1)!}(0.5)^1(1-0.5)^{4-1} = 0.25$$



Binom Dağılımlarının Şekli



Hipergeometrik Olasılk Dağılımları

- Hipergeometrik dağılımlar binom dağılımlarına büyük bir benzerlik göstermekte
- Deney n adet denemeden oluşmakta
- Deneyin olası 2 sonuçu var
- Hipergeometrik ve binom olasılık dağılmları arasındaki temel fark, hipergeometrik dağılımlarında denemeler birbirinden bağımsız değil

Hipergeometrik Olasılık Dağılımları

- Elverişli halin olasılığı bir deneyden diğerine değişiyorsa, binom olasılık dağılımları kullanılmaz. Bunun yerine hipergeometrik olasılık dağılımları uygulanılır.
  - Eğer örnek sınırlı bir popülasyondan yerine konmadan seçiliyorsa
  - Eğer örnek sayısı toplam popülasyonun % 5'inden fazla ise belli bir elverişli yada elverişsiz halin olasılık dağılımını belirlemek için hipergeometik olasılık dağılımları kullanılır.

$$P(x) = \frac{({}_{s}C_{x})({}_{N-S}C_{n-x})}{{}_{N}C_{n}}$$

N: popülasyonun büyüklüğü
S: Popülasyondaki elverişli hallerin sayısı
x: örnekteki ilgili elverişli hallerin sayısı
n: örnek yada deney sayısı
C: Kombinasyon sembolü

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

x adet nesne n sayida bir gruptan seciliyorsa

Hipergeometrik Olasılık Dağılımlarının Ortalması ve Varyansı

$$\mu = \sum xP(x) = \frac{nS}{N}$$

$$\sigma^2 = \sum x^2P(x) - \mu^2 = \frac{S(N - S)n(N - n)}{N^2(N - 1)}$$

$$= \left[n\left(\frac{S}{N}\right)\left(1 - \frac{S}{N}\right)\right]\left(\frac{N - n}{N - 1}\right)$$

Örnek: Hipergeometrik olasılık dağılımları

- Bir haftada 50 adet cep telefonu üretildiğini varsayalım. Bunlardan 40 tanesi mükemmel şekilde çalışırken, 10 tanesinde en az 1 bozuk ürün vardır. 5 örnek rastgele çekilirse, bunlardan 4'ünün problemsiz çalışma olasılığı nedir. Çekilen örnekler yerine tekrar konulmamaktadır.

$$P(4) = \frac{({}_{40}C_4)({}_{50-40}C_{5-4})}{{}_{50}C_5} = \frac{(\frac{40!}{4!36!})(\frac{10!}{1!9!})}{\frac{50!}{5!45!}} = 0.431$$

Hipergeometrik Olasılık Dağılımları

- Seçilen örnek popülasyona geri konulmuyorsa, ve örnek sayısı popülasyonun %5'inden az ise binom olasılık dağılımları hipergeometrik olasılık dağılımları yerine kullanılabilir.

Poisson Olasılık Dağılımları

- Poisson dağılımları belirli bir bölgede yada belirli bir zaman aralığında bir olayın kaç defa meydana geldiğini belirlemek açısından faydalıdır.
- Poisson dağılımlarının uygulanması için gerekli şartlar
  - Herhangi bir zaman aralığında meydana gelen elverişli hallerin sayısı diğer zaman aralıklarında meydana gelen elverişli hal sayısından bağımsızdır.
  - Bir zaman aralığındaki elverişli bir halin olasılığı, zaman aralığının büyüklüğü ile orantılıdır.
  - Olaylar ölçüm aralığında kesinlikle aynı noktada meydana gelemez.
- Örnek:
  - şirketin faks makinasına günlük gelen faks sayısı
  - İstanbulda aylık işlenen araba hırsızlığı vakası sayısı
  - Numune hastanesi acil servisine saatlik gelen hasta sayısı

Poisson Olasılık Dağılımları

$$P(x) = \frac{\mu^x e^{-\mu}}{x!}$$

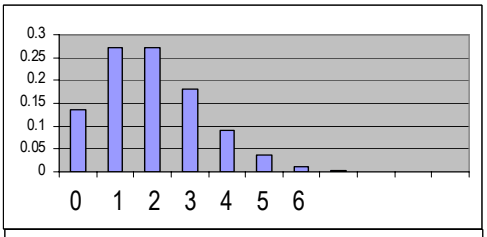
μ= ortalama elverişli hal sayısı
e=sabit sayı=2.7183
x= belli bir zaman aralığındaki elverişli hallerin sayısı (0,1,2…)

$\mu = \sum x_i\, p(x_i)$ $\sigma^2 = \sum x^2 P(x) - \mu^2 = \mu$

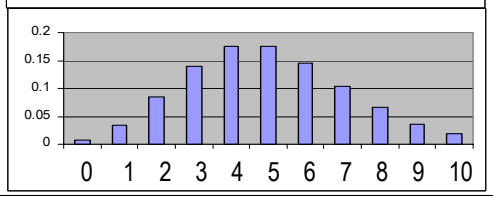
Poisson dağılımının ortalaması μ, standard sapaması ise μ'nün karaköküdür.Herhangi bir μ değerinin dağılımı pozitif bir çarpıklığa sahiptir. μ artıkça dağılım normal dağılıma yaklaşır.

Poisson Olasılık Dağılımları

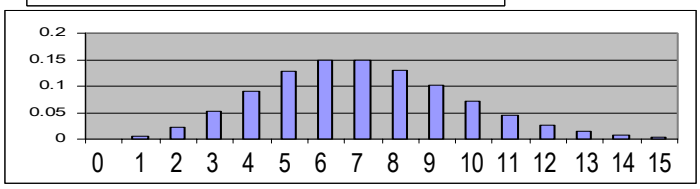
Poisson olasılık dağılımı μ =2



Poisson olasılık dağılımı μ =5



Poisson olasılık dağılımı μ =7



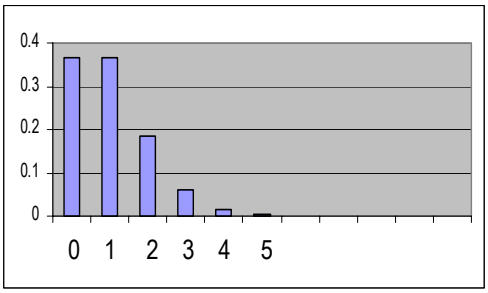
Örnek: Poisson Dağılımları

- Shell benzin istasyonuna her 15 dakikada bir ortalama 3 araç gelmektedir. Geleçek 15 dakika içinde 2 araçın gelme olasılığı nedir.

$$P(2) = \frac{3^2 (2.7183)^{-3}}{2!} = 0.224$$

Poisson dağılımları elverişli hal olasılığı küçük, np<5, ve deney sayısı(n>100)büyük olduğu durumlarda binom olasılık dağılımı yerine kullanılabilir.

Poisson Olasılık Dağılımları



$$P(X=0)=p(0)=\frac{e^{-1}1^{0}}{0!}=e^{-1}=.3678 \qquad P(X=2)=p(2)=\frac{e^{-1}1^{2}}{2!}=\frac{e^{-1}}{2}=.1839$$

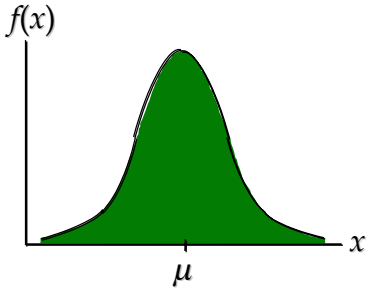
$$P(X=1)=p(1)=\frac{e^{-1}1^{1}}{1!}=e^{-1}=.3678 \qquad P(X=3)=p(3)=\frac{e^{-1}1^{3}}{3!}=\frac{e^{-1}}{6}=.0613$$

Sürekli Olasılık Dağılımları

Uniform Olasılık Dağılımı

Normal Olasılık Dağılımı

Exponent Olasılık Dağılımı



Uniform Olasılık Dağılımı

Bir random değişken olasılığı nezaman random değişkenin değer aralığı ile orantılı ise o random değişken uniform bir dağılım sergiler denir.

- Uniform Olasılık fonksiyonu

$$f(x) = 1/(b - a) \;,\; a \leq x \leq b$$
$$= 0 \quad \text{başka yerde}$$

$E(x) = (a + b)/2$

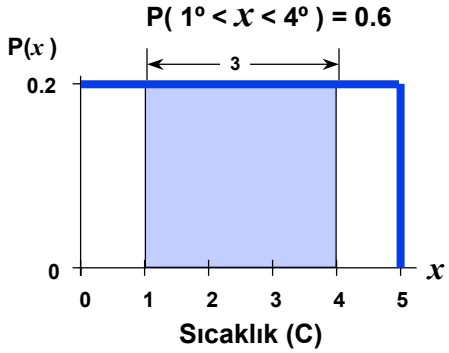
$Var(x) = (b - a)^2/12$

a = bir değişkenin varsayacağı en küçük değer

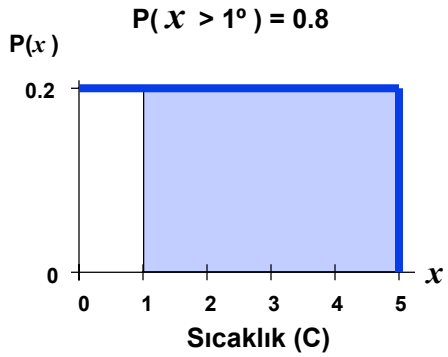
b = bir değişkenin varsayacağı en büyük değer

Uniform Olasılık Dağılımları

Taralı alan = 3 • 0.2 = 0.6



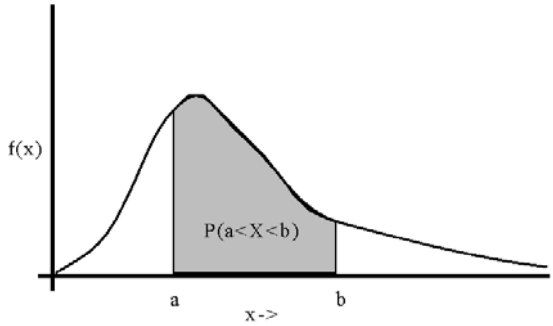
Şekil 1



Şekil 2

Sürekli Olasılık Dağılımları

- Sürekli bir random değişkenin(X) olasılığı (değişkenin değerinin a-b aralığında bir değer aldığını varsayarsak), a-b aralığında olasılık dağılım fonksiyonu eğrisi altında kalan alan olarak tanımlanır.



$$f(x) \geq 0$$

$$\int_{x=-\infty}^{x=+\infty} f(x)dx = 1$$

$$P(a < X < b) = \int_{x=a}^{x=b} f(x)dx$$

Ortalama ve Varyans

- ✓ Sürekli bir random değişkenin aritmetik ortalaması,

$$E(X) = \int_{-\infty}^{+\infty} f(x)\,dx$$

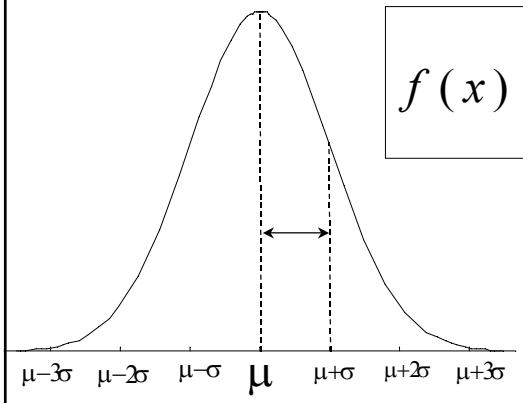
- ✓ Varyansı

$$Var(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x)\,dx$$

Normal Olasılık Dağılımı

- Doğadaki, endüstrideki bir çok olay normal veya normala yakın bir dağılım göstermektedir.

Normal olasılık dağılım eğrisinin fonksiyonu aşağıdaki şekilde tanımlanır.



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

μ = A. Ortalama,

σ = standart sapma

π = 3.14159, e = 2.71828

Normal Olasılık Dağılım

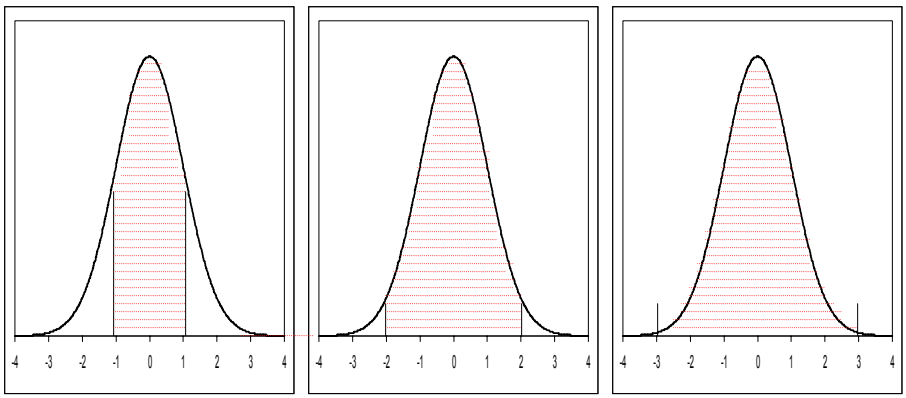
- Özellikleri
  - Çan eğrisi şekilli bir dağılım
  - $-\infty \leq x \leq +\infty$
  - Normal olasılık dağılım eğrisinin altında kalan alan 1'e eşittir.

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

  - Normal dağılımlar μ etrafında simetriktir.
  - μ Normal dağılımının yerini belirlemektedir ve grafikteki en yüksek noktadır.
  - Normal dağılımlarda aritmetik ortalama, mod ve medyan birbirine eşittir.
  - σ yayılma miktarını belirlemektedir.

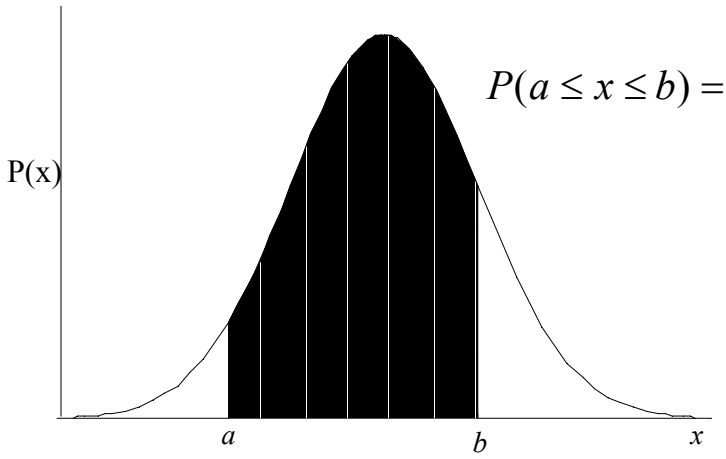
Normal Olasılık Dağılımı

- Normal dağılımlarlarda
- $\mu \pm 1\sigma$ altında kalan alan tüm dağılımın % 68 ini
- $\mu \pm 2\sigma$ altında kalan alan tüm dağılımın % 95 ini
- $\mu \pm 3\sigma$ altında kalan alan tüm dağılımın % 99.7 ini kapsamaktadır.



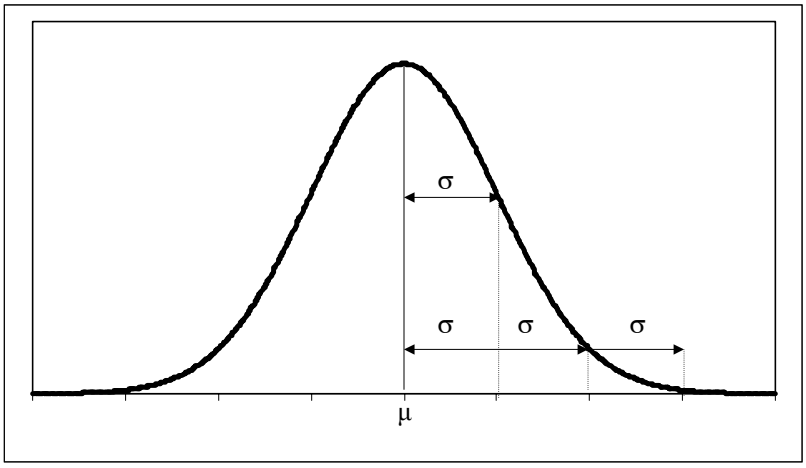
$\mu \pm 1\sigma$ 68% | $\mu \pm 2\sigma$ 95% | $\mu \pm 3\sigma$ 99.7%

Normal bir olasılık dağılımında olasılıklarının gösterimi



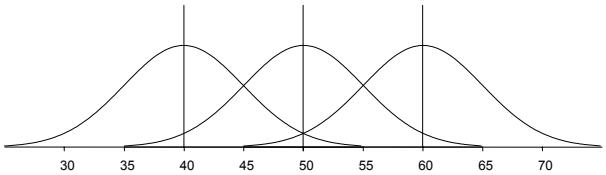
$$P(a \le x \le b) = \int_a^b f(x)dx$$

Normal Dağılım

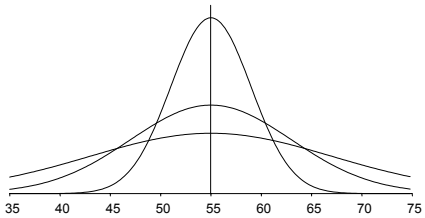


Sonsuz sayıda normal olasılık dağılımları vardır.

Şekil 1. Farklı A. ortalamalı, aynı standart sapmalı normal olasılık dağ.

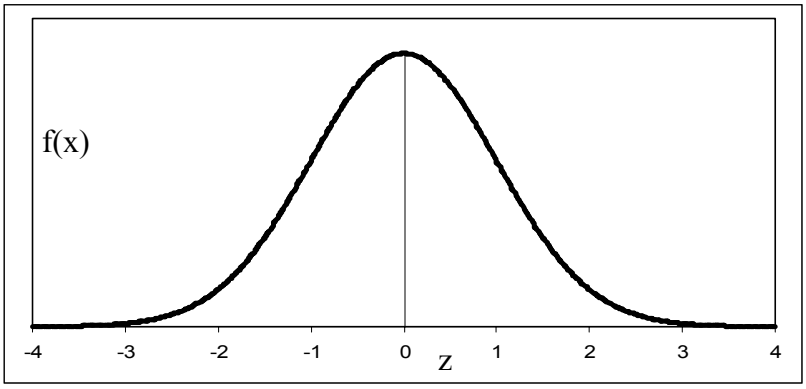


Şekil 2. Aynı A. ortalamalı farklı standart sapmalı normal olasılık dağ.



Standart Normal Dağılım

- Normal dağılımlarının tümü standart normal dağılımla ilişkilidir.
- Standart normal değişken z'nin sergilediği A. ortalaması 0, varyansı 1'e eşit olan normal dağılıma **standart normal dağılım** denir.



Normal Dağılım

- Her bir normal random değişken (x), aşağıdaki formül yardımıyla standart normal değişkene (z) dönüştürülür.

$$z = (x - \mu)/\sigma$$

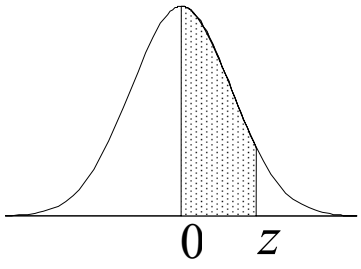
- Standard normal dağılım tablo halin.

Standart Normal Dağılımın Özellikleri

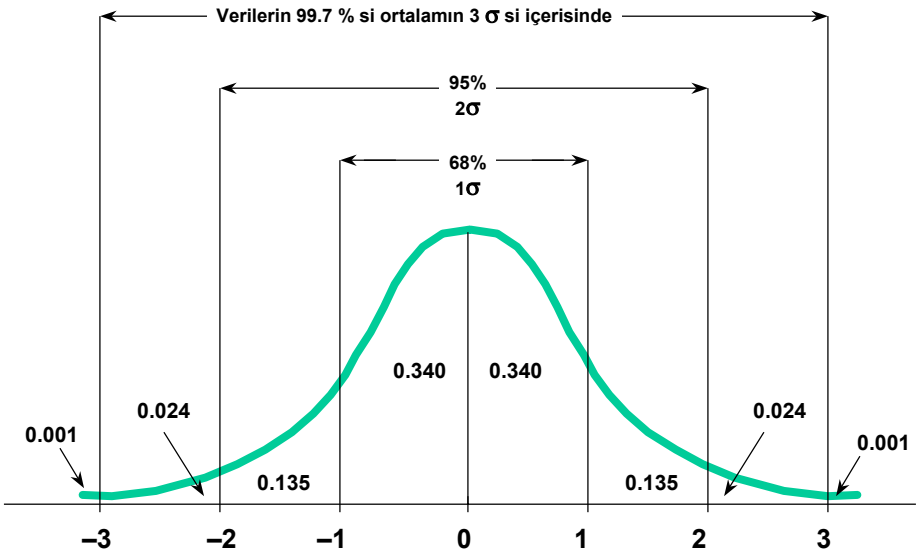
•Normal eğrinin altında kalan alan 1'e eşittir.

•Dağılım A. ortalama etrafında simetriktir ve her iki yönde yatay eksene yaklalarak fakat dokunmadan devam etmektedir.

- Aritmetik ortalama 0, eğriyi iki eşit parcaya (0.5) bölmektedir.
- Hemen hemen tüm alan
- z = -3.00 ve z = 3.00 değerleri arasındadır.



Standard Normal Dağılım : $\mu = 0$ ve $\sigma = 1$



Standard Normal (z) Dağılım Tablosu

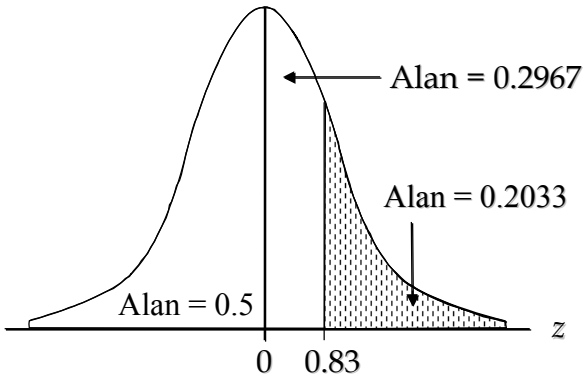
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Standart Normal Olasılık Dağılımı

P(z<0)

P(0<z<0.83)

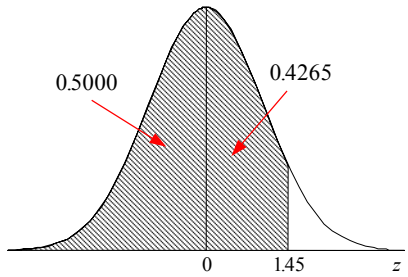
P(z>0.83)



Standart Normal Dağılım Tablosunun Kullanımı

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389

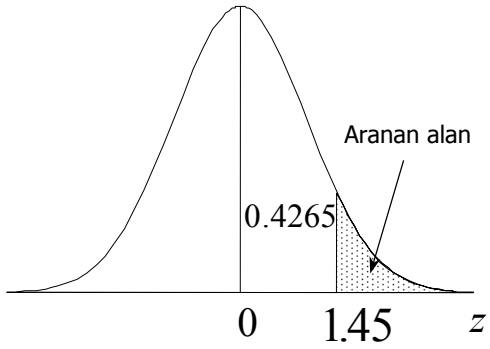
Örnek 2 (Tablo methodu)



$$P(z < 1.45) = 0.5000 + 0.4265 = 0.9265$$

Örnek 3

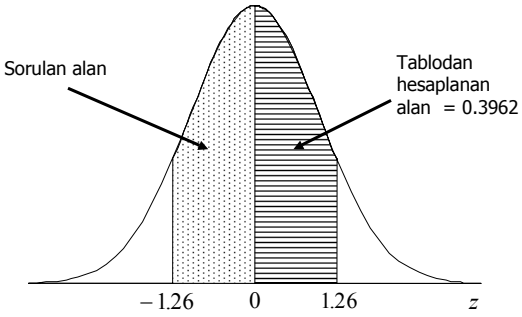
- P(z > 1.45)=?



$$P(z > 1.45) = 0.5000 - 0.4265 = 0.0735$$

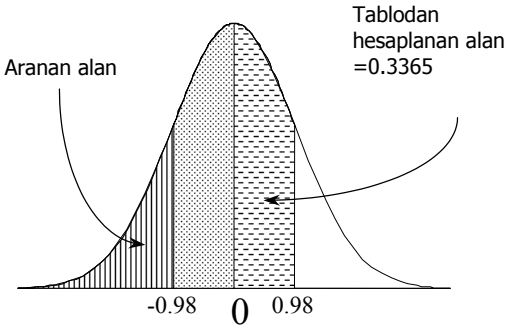
Örnek 4

- P(-1.26 < z < 0) =?
- = P(0 < z < 1.26)
- = 0.3962



Örnek 5

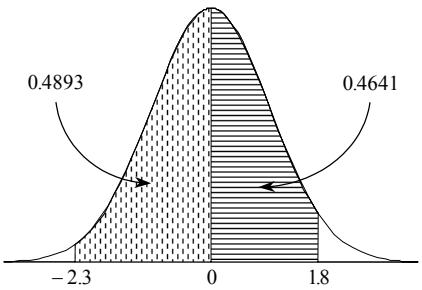
- P(z < -0.98)=?



$$P(z < -0.98) = 0.5000 - 0.3365 = 0.1635$$

Örnek 6

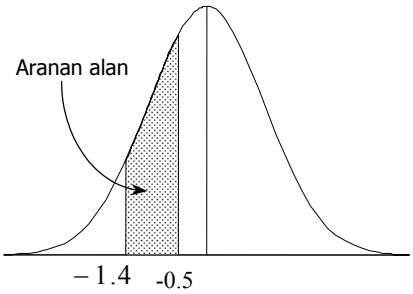
- P(-2.3<z<1.8)=?



$$P(-2.3 < z < 1.8) = P(-2.3 < z < 0) + P(0 < z < 1.8)$$
$$= 0.4893 + 0.4641 = 0.9534$$

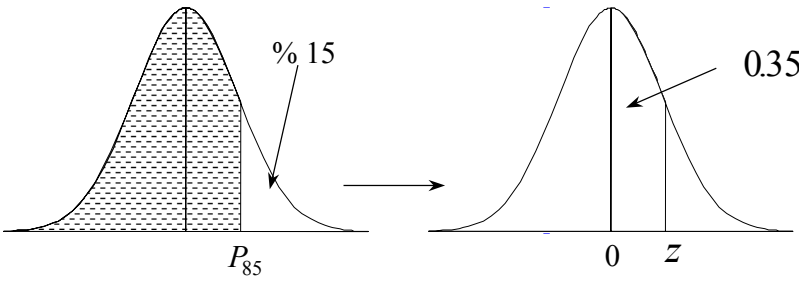
Örnek 7

- P(-1.4<z<-0.5)=?



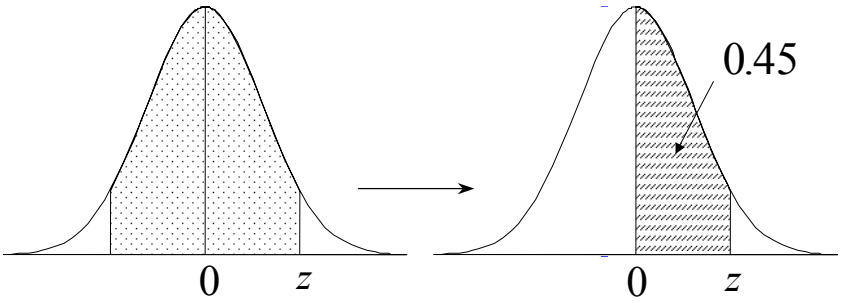
$$P(-1.4 < z < -0.5) = P(0 < z < 1.4) - P(0 < z < 0.5)$$
$$= 0.4192 - 0.1915 = 0.2277$$

- Eğer eğrinin altında kalan alan biliniyor ise, standart normal dağılım tabloları z değerini belirlemek için kullanabilir.



z	0.00	0.01	0.02	0.03		0.04	0.05
⋮							
1.0				0.3485	0.3500	0.3508	
⋮							

- Aşağıdaki standart normal eğride % 90'lık alana karşılık gelen z değeri bulalım.

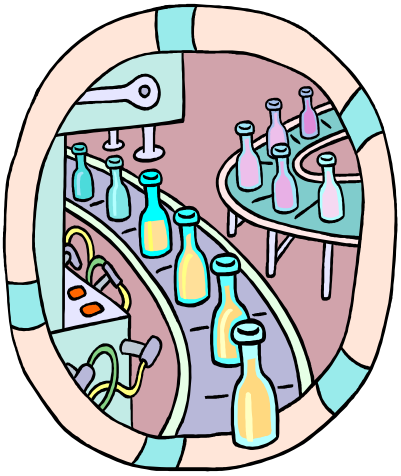


z	0.00	0.01	0.02	0.03	0.04		0.05
.							
1.6					0.4495	0.4500	0.4505
.							

Örnek 8

•Bir şişeleme makinası ortalama olarak 32 ml sodayı 0.02 ml standart sapmayla doldurucak şekilde ayarlanmıştır.

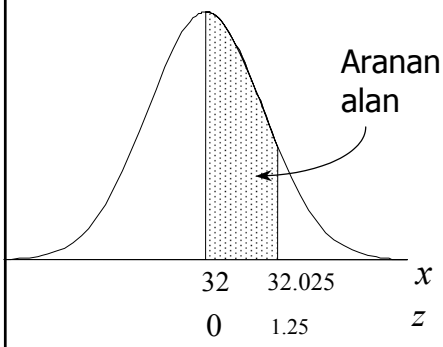
•Dolum miktarının normal bir dağılım gösterdiğini varsayarsak, rastgele seçtiğimiz bir şişenin 32-32.025 ml arasında soda içerme olasılığı nedir?



Çözüm 8

$$x = 32; \quad z = \frac{32-\mu}{\sigma} = \frac{32-32}{.02} = 0$$

$$x = 32.025; \quad z = \frac{32.025-\mu}{\sigma} = \frac{32.025-32}{.02} = 1.25$$

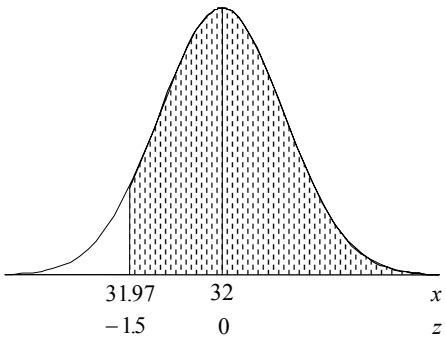


$$P(32 < x < 32.025) = P\left(\frac{32-32}{.02} < \frac{x-32}{.02} < \frac{32.025-32}{.02}\right)$$

$$= P(0 < z < 1.25) = 0.3944$$

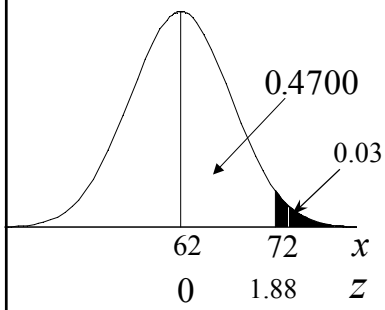
•Rastgele seçtiğimiz bir şişenin 31.97 ml den fazla soda içerme olasılığı nedir?

$$P(x > 31.97) = P\left(\frac{x-32}{.02} > \frac{31.97-32}{.02}\right) = P(z > -1.5)$$
$$= 0.5000 + 0.4332 = 0.9332$$



Örnek 9

•Bir polis radarı akşam trafiğinde E-5de araçlarının hızları denetlemektir. Araçlarının hızları aritmetik ortalması 62 km/saat olan normal bir dağılım göstermektedir. Araçlarının %3 ü 72 km/saat 'in üzerinde seyahat ediyorsa tüm araçlarının hızlarının standart sapmasını hesaplayınız?



$$P(x > 72) = 0.03$$

$$P\left(\frac{x-62}{\sigma} > \frac{72-62}{\sigma}\right) = 0.03$$

$$P\left(z > \frac{72-62}{\sigma}\right) = 0.03 \qquad P(z > 1.88) = 0.03$$

$$\frac{72-62}{\sigma} = 1.88$$

$$(1.88)(\sigma) = 10$$

$$\sigma = 10/1.88 = 5.32$$

Binom Olasılık Dağılımlarının Normal Dağılıma yaklaşmas Kural

- Normal dağılım tablolarının binom dağılım yerine kullanabilmesi için

$$(np > 5) \quad \& \quad n(1-p) > 5$$

Standart Normal Dağılım

Bir standart normal dağılımda aşağıdaki koşulları sağlayan k değerlerini bulalım.

(a) $P(z < k) = 0.1271$

(b) $P(z < k) = 0.9495$

(c) $P(z > k) = 0.8186$

(d) $P(z > k) = 0.0073$

(e) $P(0.90 < z < k) = 0.1806$

(f) $P(k < z < 1.02) = 0.1464$

Exponent Olasılık Dağılımı

- Exponent Olasılık Fonksiyonu

$$\boxed{f(x) = \frac{1}{\mu} e^{-x/\mu}}$$

$x \geq 0,\ \mu > 0$

μ = A. ortalama

e = 2.71828

ÖRNEKLEME TEORİSİ

Bir popülasyonun istatistiksel parametrelerini belirlemede örneklemenin tercih edilmesinin bir çok sebebi vardır. Bunlar şu şekilde özetlenebilir.

1. Çok sıkca, seçilen örneklerin üyelerinin yok edilmesi durumunda, örneklerinin popülasyona geri katılamaması
2. Popülasyonun tümüne ulaşmanın örneklemede mümkün olmayabilmesi
3. Popülasyonun tümünü örneklemenin maliyetinin yüksek olması
4. Doğru seçilmiş bir örneklemenin popülasyonun parametrelerini uygun bir şekilde tahmin edebilmesi ve bunun sonuçunda maliyet ve zaman kaybının azaltılması
5. Popülasyonun tüm üyeleriyle bağlantı kurmanın çok zaman alması

2 türlü örnekleme vardır: Rastgele ve karara dayalı yani iradi örnekleme

Rastgele örneklemenin bir kaç türü vardır:

1. **Basit rastgele örneklemede** popülasyonun her bir üyesi aynı şeçilme şansına yada olasılığına sahiptir.
2. **Sistematik rastgele örneklemede** rastgele bir başlangıç noktası seçilir ve ondan sonra her n'inci popülasyon üyesi örneklemede seçilir.
3. **Cluster örneklemede** popülasyon gruplara ayrılır ve bu gruplardan rastgele örnekleme yapılır.

Karara dayalı yani iradi örneklemede örnek şeçimi tamamiyle örneklemeyi yapan kişinin kararına dayalıdır. Dolayısıyla bu tür örneklemede popülasyon parametrelerinin tahmininde hataya sebeb olabilir.

Popülasyon parametresi ile örnekleme istatistiği arasındaki arasındaki fark **örnekleme hatası** olarak tanımlanır.

Bir örneklemeden tahmin edilen ortalamadaki standart hata miktarı :

$$\sigma_{\overline{X}} = \frac{s}{\sqrt{n}}$$

s: örneklemedeki gözlemlerin standart sapması
n: örneklemedeki gözlemlerin sayısı

Merkezi limit teoremine göre eğer bir popülasyon normal bir dağılım gösteriyorsa örneklemelerinin aritmetik ortalamalarının dağılımıda ayrıça normal bir dağılım gösterir. Eğer popülasyon normal dağılım sergilemiyorsa örneklemelerinin aritmetik ortalamalarının dağılımı örnek sayısı artıkça normale yakın bir dağılım gösterir.

Popülasyonun parametrelerinin belirlenmesinde 2 türlü tahminden yararlanılır:

Nokta tahmini(point estimate): popülasyon parametresinin tahmininmde tek bir değer kullanılır.
Enterval yada aralık tahmini (interval estimate): popülasyon parametresinin hangi değerler arasında bulunacağının belirlenmesidir.

Aralık tahmininde popülasyon parametersinin güven aralığının hesaplanması gerekmektedir. Bir popülasyonun ortalamasının güven aralığı örneklemedeki gözlem sayısına (n), örneğin standard sapmasına, ve güven aralığının derecesine bağlıdır.

Bir ortalamanın güven aralığı şu şekilde genel olarak ifade edilebilir:

$$\overline{X} \pm z\sigma_{\bar{x}}$$

$$\overline{X} \pm z\frac{s}{\sqrt{n}}$$

s: örneklemenin standard sapması
z: standard değer
n: örneklemedeki gözlem sayısı

%95 ve % 99 güven aralığı istatiksel tahminlerden en yaygın olarak kullanılmaktadır.

n ≥ 30 için

%95 güven aralığı:

$$\overline{X} \pm 1{,}96\frac{s}{\sqrt{n}}$$

%99 güven aralığı:

$$\overline{X} \pm 2{,}58\frac{s}{\sqrt{n}}$$

1,96 ve 2,58 değerleri gözlemlerinin sırasıyla %95 ve %99 una karşılık gelen standard değerlerdir. Bu güven aralıklarına karşılık gelen değerler standart normal dağılım tablolarından hesaplanır. Örneğin Bu tablo yarım normal dağılıma göre

hazırlandığından 0,95/2 =0,475 . Bu değere karşılık gelen standart değer normal dağılım tablosundan 1,96 olarak kolaylıkla okunabilir.

Bir örneklemede popülasyonun ortalamanın belirlenmesi için belirlenecek gözlem sayısı(n), seçilecek güven aralığına(z), izin verilebilir maksimum hata oranına(E), ve verilerin standard sapmasına(s) bağlıdır.

$$n = \left(\frac{z.s}{E} \right)^2$$

Eger örneklemediki veri sayısı(n) tüm popülasyonun (N) %5 inden büyük ise yani n/N > 0.05 ise hem popülasyon ortalamasının hemde oranının standard hata miktarına bir düzeltme uygulamak gerekmektedir. Bu düzeltme katsayısı şu şekilde ifade edilir.

$$\frac{N-n}{N-1}$$

Popülasyon ortalaması için standart hataya uygulanaçak düzeltme (n/N > 0.05)

$$\sigma_{\overline{X}} = \frac{s}{\sqrt{n}} \left(\frac{N-n}{N-1} \right)$$

aynı düzeltmeyi ortalamanın güven aralığı için yazarsak,

$$\overline{X} \pm z \frac{s}{\sqrt{n}} \left(\frac{N-n}{N-1} \right)$$

Örneklemedeki gözlem sayısının popülasyonun tümüne oranı % 5 den az ise düzeltme katsayısının standart hataya katkı payı çok küçüktür o nedenle önemsenmeyebilir. Aksi durumda (n/N >%5) düzeltme miktarı standart hatayı azaltaçağından popülasyon ortalamasının aralığı daralaçaktır. Buda doğaldır çünkü örnek sayısı artıkça ortalamın standart hatasıda doğal olarak azalma göstereçektir.

Hipotez Testleri

Dr. İrfan Yolcubal
Kocaeli Üniversitesi
Jeoloji Müh.

Hipotez

- Örneklemeye dayalı bir popülasyon parametesinin değeri hakkında ileri sunulan iddia

Örnekler:

1. İstatistik Vize sınıvanın ortalaması 50'nin altındadır.
2. Televizyon izleyicilerin %70 i günlük haber programlarını izlemektedir.
3. Firestone ve Lassa tatafından üretilen lastiklerinin ömrü aynıdır.

Hipotez Testleri

- Bir popülasyon hakkında ileri sunulan hipotezinin kabul edilip edilmeyeceğini belirlemek için örneklemeye dayalı sistematik izlenen bir seri işlemler.

  5 aşamadan oluşur.

1. Null ve alternatif hipotezin belirlenmesi

Null hipotezi: Bir popülasyon parametresi hakkında ileri sürülen varsayım. Genellikle bu varsayımda popülasyon parametresinin belli bir değeri olduğu varsayılır.

- H_0= null hipotezi yada sıfır hipotez

Alternatif hipotez: Örneklemeye ait veriler null hipotezonin yanlış olduğuna ait deliller sunduğu durumlarda kabul edilen hipotezdir

- H_A = alternatif hipotez

Hipotez Testinin Aşamaları

2. Önem veya Risk Derecesinin Belirlenmesi(α): Aslında doğru olan Null hipotezinin rededilme olasılığı:

Risk derecesinin seçimi tercihe dayalı

- Genelde 0.05 yani % 5 ve % 1 risk dereceleri araştırmalarda kullanılmakta

Hata Tipleri

- **I. tip hata**: Null hipotezi doğru iken reddedilir.
- I. Tip hata yapma olasılığı α olarak bilinmektedir.
- **II. tip hata**: Null hipotezi yanlış iken rededilmez.
- II. Tip hata yapma olasılığı β olarak bilinmektedir.
- Daima bu hatalardan birini yapma ihtimali vardır. Bu ihtimalleri risk derecesini belirleyerek azaltmak isteriz.

Hipotez Testlerinde Hatalar

Karar	Null Hipotezi doğru	Null Hipotezi Yanlış
Null hipotezi Kabul etme	Doğru Karar	I. tip hata
Null hipotezi redetme	II. tip hata	OK

Hipotez Testinin Aşamaları

3. Istatistiksel test metodunun belirlenmesi: Null hipotezin rededilip edilmeyeceğinin belirlenmesinde kullanılan ve popülasyon örneklemesinden elde edilen değer

örnek: t, F,ve ki kare istatistik testleri

4**. Null hipotezinin hangi koşullarda kabul ve hangi koşullarda rededileceğinin belirlenmesi**

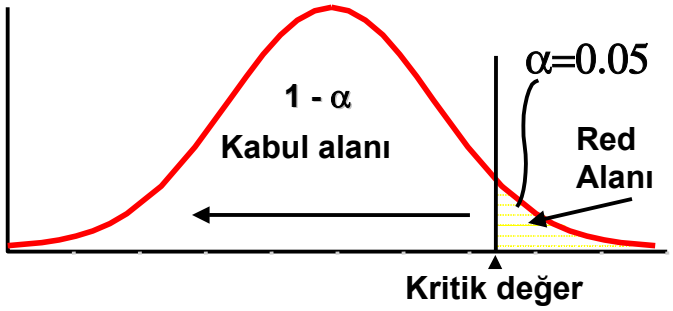
5. Karar verilmesi: Null hipotezinin alınan risk derecesi doğrultusunda reddi yada kabülü.

1: Null ve Alternatif hipotezleri ileri sürmek

- Farzedelim ögrençilerin ders geçmek için 60 almaları gerekmekte.
- Rastgele 40 ögrenci secelim ve onların ortalamalarının 64 olduğunu varsayalım
- Araştırma sorusu: Popülasyonun gerçek ortalaması 60 ın üzerinde midir?
  - H_0: $\mu \leq 60$
  - H_A: $\mu > 60$

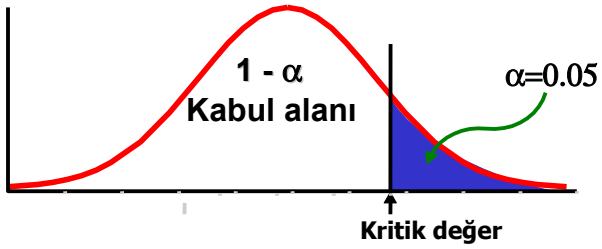
2: Önem Derecesini(α) belirlemek

- Önem derecesi, null hipotezi gerçekten doğru iken, null hipotezini redetme olasılığıdır
- Örnek: α = 0.05 seçelim



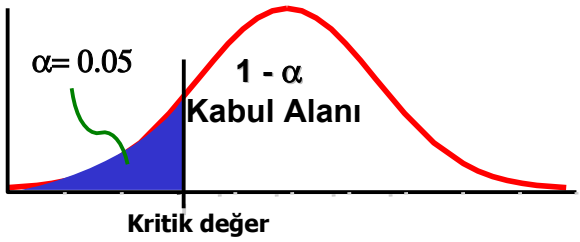
3: Hipotez testinin 1 veya 2 yönlü olup olmadığının belirlenmesi

- Eğer alternatif hipotez ortalamanın belli bir değere eşit yada ondan büyük olduğunu ifade ediyor ise hipotez tek yönlüdür.
- Örnek: H_A: $\mu \geq 60$



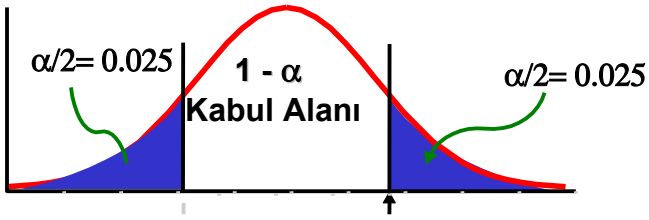
3: (Devam)

- Eğer alternatif hipotez ortalamanın belli bir değere eşit yada ondan küçük olduğunu ifade ediyorsa, hipotez tek yönlüdür. Örnek: H_A: $\mu \leq 60$



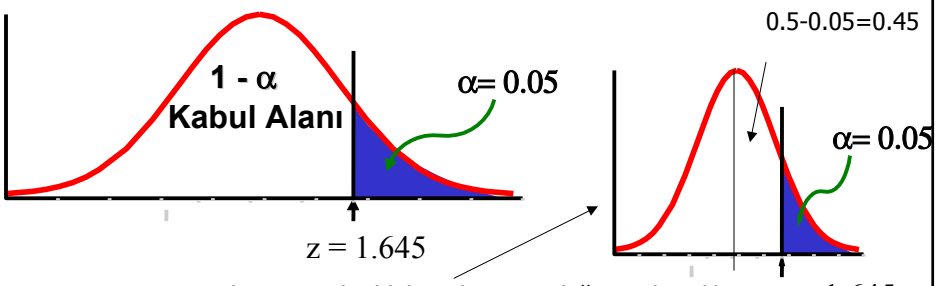
3: (Devam)

- Eğer Alternatif hipotez ortalamanın belli bir değere eşit olmadığını ifade ediyor ise bu hipotez çift yönlüdür. H_A: $\mu \neq 60$



4: Kritik değeri veya değeleri belirlemek

- Bilinmek istenilen – Null hipotezi doğru varsayarsak dağılımın 1-α yüzdesine karşılık gelen kritik değer.
- Eğer popülasyonun stadart sapması (σ) biliniyor ise yada σ bilinmiyor fakat $n \geq 30$ ise standart normal tabloları kullanılarak risk derecesine karşılık gelen z kritik değeri belirlenir.



Standart normal tablolarından 0.45 değerine karşılık Gelen z değeri bulunur. z=1.645

5: Test istatistiğini belirlemek ve kritik değerle karşılaştırmak

Popülasyonun standart sapması bilniyor ise z= kritik değer sağdaki formül vasıtasıyla hesaplanır.

$$z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

$\overline{X}$ = orneklemenin ortalamasi

μ = populasyon ortalamasi

σ = populasyonun standart sapmasi

- Popülasyonun standart sapması bilinmiyorsa ve $n \geq 30$, örneklemenin standart sapması (s) popülasyonun standart sapması yerine kullanılabilir.
- Populasyon normal dağılım sergilemekte
- Hipotez testinde kullanıcak değer:

$$z = \frac{\overline{X} - \mu}{s / \sqrt{n}}$$

Hipotez Test Aşamalarını Özetlersek

1. Null ve Alternatif Hipotezleri Belirlemek: H_0, H_a
2. Önem yada Risk Derecesini Belirlemek: α
3. Hipotezin tek mi çift mi yönlü olduğunu belirlemek
4. Kritik değerleri belirlemek
5. Test istatistik değerlerini hesaplamak ve kritik değerle karşılaştırmak

Örnek 1

$$H_0: \mu = 50$$
$$H_1: \mu \neq 50$$

Örnek ortalaması 49, örneklemedeki veri sayısı da 36dır. Popülasyonun standart sapması ise 5 dir. Hipotez testinde % 5 risk alırsak

a) Hipotez testi tek mi yoksa çift mi yönlüdür
b) Null hipotezi hakkındaki kararınız nedir
c) Bu kararı almaktda nekadar kendinize güveniyorsunuz yani *p* değeri nedir.

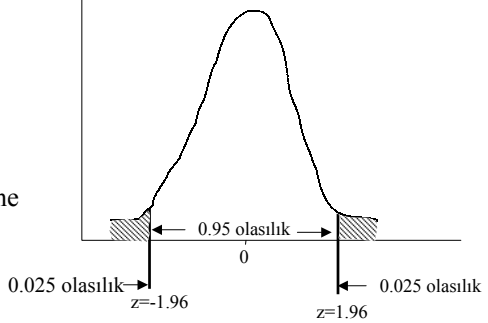
Örnek 1. Çözüm

a) Hipotez testi iki taraflı bir hipotezdir çünkü alternatif hipotezin yönü yoktur yada belli değildir. Popülasyon ortalaması 50 den farklı olabilir ifadesi büyükte olabilir ve küçükte olabilir gibi 2 ihtimal içermektedir. Bu nedenle hipoteze 2 taraflı hipotez denilmektedir.

b) %5 riskle taralı alanlar hipotezin rededildiği alanları ifade etmektedir.

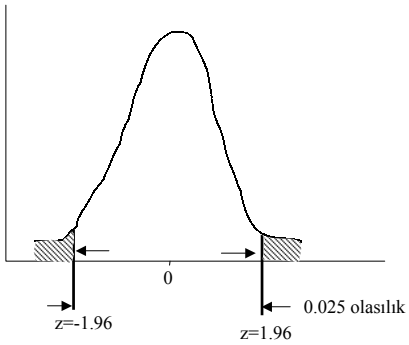
$$z = \frac{49 - 50}{5/\sqrt{36}} = -1.2$$

Hesaplanan z değeri bu taralı alanlar dışında kalan bölgeye düştüğüne göre Null hipotezini kabul edebiliriz



Örnek 1. Çözüm (Devam)

c) Null hipotezini kabul etmede ne kadar eminiz ? Bunu belirleye bilmek için hesaplanan z değerinin o değerinin üzerinde bulunma olasılığını yani p değerini hesaplamamız gerekecektir.



–1.2 ve altında bir değer olma olasılığı 0,1151dir (0.5-0.3849). *p* değerini hesaplayabilmek için z değerinin –1.2 den az ve 1.2 den fazla olma ihtimalini hesaplamamız gerekmektedir çünkü hipotez iki taraflı olup iki farklı red bölgesi içermektedir. Bu nedenle *p* değeri 2 x 0,1151di r. *p* değeri risk derecesinden 0.05 büyük olduğundan null hipotezi kabul edilir.
p değeri popülasyonun ortalmasının 50 nin üzerinde veya altında olma olasılığının %11.51 olduğunu ifade eder.

Örnek 2: Tek yönlü z testi

- Bir kutu mısır gevreği 368 gramın üzerinde midir?
- Ratgele seçilen 25 kutunun ortalaması $\bar{X}$ = 372.5 gr.
- Üretiçi firma ürün miktarı için standart sapmayı σ 15 gram olarak belirlemiştir.
- Hipotezi 0.05 önem derecesi ile test edelim.

Tek yönlü hipotez test çözümü

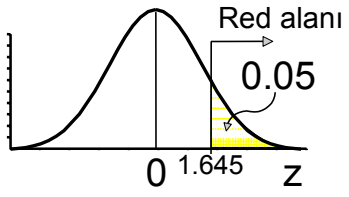
H_0: $\mu \leq 368$

H_A: $\mu > 368$

$\alpha = 0.05$

$n = 25$, σ bilinmekte

Kritik değerler



Test İstatistiği:

$$z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{372.5 - 368}{\frac{15}{\sqrt{25}}} = +1.50$$

Karar:

Null hipotez $\alpha = 0.05$ ile rededilmez

Sonuç:

Ortalamanın 368 gr üzerinde olduğuna ait yeterli delil yoktur.

Çift yönlü z Testi Çözümü

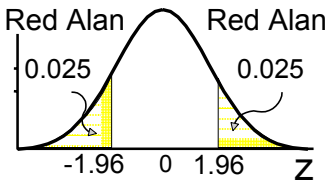
H_0: $\mu = 368$

H_A: $\mu \neq 368$

$\alpha = 0.05$

$n = 25$, σ bilinmekte

Kritik değerler



Test İstatistiği:

$$z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{3725 - 368}{\frac{15}{\sqrt{25}}} = +1.50$$

Karar:

Null hipotezi $\alpha = 0.05$ ile red edilmez

Sonuç:

Ortalama miktarın 368 olduğu hakkında yeterli bir delil yoktur

ÖDEV

$$H_0: \mu \leq 10$$
$$H_1: \mu > 10$$

Örnek ortalaması 12, örneklemedeki veri sayısı da 36dır. Popülasyonun standart sapması ise 3 dir. Hipotez testinde % 2 risk alırsak

a) Hipotez testi tek mi yoksa çift mi yönlüdür
b) Null hipotezi hakkındaki kararınız nedir
c) Bu kararı almaktda ne kadar kendinize güveniyorsunuz?

HİPOTEZ TESTİ:İKİ POPÜLASYONUN ORTALAMALARININ KARŞILAŞTIRILMASI

Amaç:2 örnek ortalamasının aynı ortalamalı 2 popülasyondan gelip gelmediğini test etmek

$$z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Hipotez test edilirken daha önceki kısımlarda bahsettiğimiz hipotezıin 5 aşamasıda aynı şekilde uygulanır. Sadece fark z değerinin hesaplanışıdır.

Örnek: 2 popülasyonun ortalamalarının karşılaştırılması

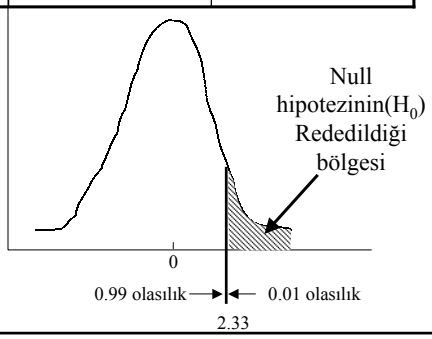
2 farklı hastanenin acil servisine gelen hastalara müdahele süresi aşağıda sunulmaktadır. Bu araştırmaya göre %1 riskle numune hastanin acil servisi, sigorta hastanesinin acil servisinden daha mı hızlı hastalara ilk müdaheleyi yapmaktadır?

Hastane	Ortalama süre	Örnek standart sapması	Örnek sayısı
numune	5.5 dak	0.4 dak	50
sigorta	5.3 dak	0.3 dak	100

Null ve alternatif hipotez:

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \rangle \mu_2$



Örnek: Devam

$$z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{5.5 - 5.3}{\sqrt{\frac{0.4^2}{50} + \frac{0.3^2}{100}}} = 3.12$$

$z = 3.12 > 2.33$

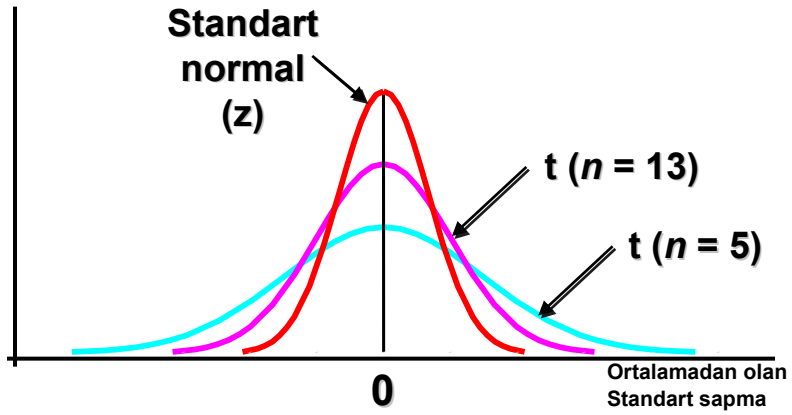
null hipotezi red edilir, alternatif hipotez %1 riskle kabul edilir.

p değeri bu büyüklükte yada onun üzerinde bir değer bulma olasılığıdır.
3.12 ve üzerinde bir z değeri alma olasılığı 0.499(Tabloda 3.12 değeri olmadığından en yakın 3.09 değerine karşılık gelen olasılık esas alınmıştır.
Buna göre 3.12 ve üzeri bir değer olma olasılığı: 0.5-0.499=0.001
Bu değer 0.01 risk derecesinden küçük olduğundan null hipotezinin doğru olmama ihtimali çok yüksektir.

z-testi and t-testi karşılaştırılması

- **z-test istatistiği**
  - **Normal dağılıma dayalı**
  - Popülasyonun varyansı bilindiğinde yada örneklemedeki veri sayısı büyük olduğunda örneklerin ortalamaları hakkındaki hipotezleri test etmek için kullanılır
- **t-test istatistiği**
  - t dağılımına dayalı
    - t dağılımının şekli örneklemedeki veri sayısına bağlı olarak değişmektedir
    - Serbestlik derecesine bağlıdır df :n-1
    - Örneklemedeki veri sayısı artıkca t dağılımı normal dağılıma yaklaşır
  - Popülasyonun varyansı yada standart sapması bilinmediğinde ve örneklemedeki veri sayısı küçük olduğunda (n<30) örneklerin ortalamaları hakkındaki hipotezleri test etmek için kullanılır

t dağılımları



t-testleri

Varyans hakkında kesin bir bilgiye sahip olmadığımız için (sadece tahmin), t dağılımını kullanırız

- t-testinin ortalaması $$t = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$$

$\overline{X}$ = örnek ortalaması

μ = test edilen popülasyonun ortalaması

s = örnek standart sapması

n = örneklemedeki veri sayisi

Hipotez testi: σ bilinmemekte

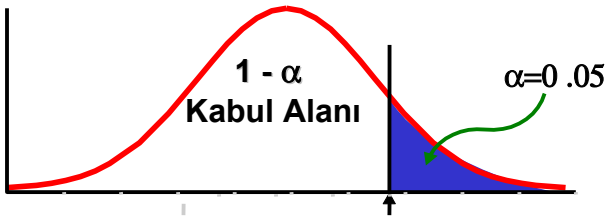
- Yüksek lisans dersindeki ögrencilerin araştırma metodları hakkında iyi bir bilgiye sahip olup olmadıklarını ögrenmek istiyorum
- 6 ögrenci sınıftan rastgele seçilir ve sınava tabii tutulur
- Sınıfın test den en az 70 alabilmesini istiyorum. 6 ögrencinin notları 62, 92, 75, 68, 83, and 95.

1 & 2: Hipotezleri belirle, önem derecesini α belirle

- Sınıfın ortalama notu 70 ve üstüdür:
  - H_0: $\mu \leq 70$
  - H_A: $\mu > 70$
- $\alpha = 0.05$

3: Tek veya çift yönlü bir hipotez mi?
4: Kritik değerleri belirle

- Tek yönlü
- Kritik değerleri t tablosundan belirlenir



t= 2.015
[serbestlik derecesi d.f. =5] için

5: Test istatistklerini hesapla & değerlendir

- t değeri kritik değerinden küçüktür. Null
- Hipotezi kabul edilir

$$\bar{x} = \frac{475}{6} = 79.17$$

$$s = 13.17$$

$$t = \frac{79.17 - 70}{\frac{13.17}{\sqrt{6}}} = 1.71$$

Örnek hakkında Sorular

- Önem derecesini 0.1 olarak seçersek ne olur?
- Kritik değer= 1.476
- Null hipotezini red et!
- Eğer α =0 .05 fakat test çift yönlü ise ne olur ?
- Bu hipotez testini örneğin büyüklüğü nasıl etkilemiştir.

Örnek: 2 yönlü t testi

- Kuzey Kıbrıstaki seçim noktalarının her birinde az yada çok 368 seçmen oy kullanmışmıdır?
- 36 rastgele seçim noktasındaki ortalama seçmen sayısı 372.5 ve standart sapma 12 seçmendir.
- Hipotezi 0.05 önem derecesi ile test edelim

Çift yönlü t Testi: Çözüm

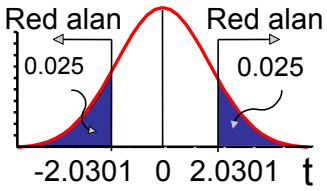
H_0: $\mu = 368$

H_1: $\mu \neq 368$

$\alpha = .05$

df = 36 - 1 = 35

Kritik değerler



Test İstatistiği:

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{372.5 - 368}{\frac{12}{\sqrt{36}}} = +2.25$$

Karar:

Null hıpotezi α = 0.05 ile red edilir

Sonuç:

Popülasyonun ortalamasının 368 olmadığına ait delil vardır

Örnek: Tek yönlü t testi

- Kuzey Kıbrıstaki seçim noktalarının her birinde 368 den fazla seçmen oy kullanmış mıdır?
- 36 rastgele seçim noktasındaki ortalama seçmen sayısı 372.5 ve standart sapma 12 seçmendir.
- Hipotezi 0.05 önem derecesi ile test edelim

Tek yönlü t testi - Çözüm

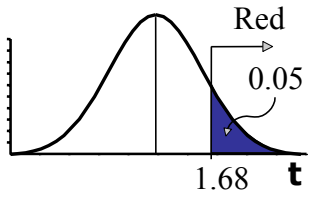
H_0: $\mu \leq 368$

H_1: $\mu > 368$

$\alpha = .05$

df = 36 - 1 = 35

Kritik değerler



Test İstatistiği:

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{372.5 - 368}{\frac{12}{\sqrt{36}}} = +2.25$$

Karar:

Null hıpotezi α = 0.05 red edilir

Sonuç:

368 den fazla seçmenin ortalama oy sandıklarında oy kullandığına ait delil vardır

t testi

- Popülasyon ortalamasının testi:bir önceki slaytlarda bahsettik
- 2 birbirinden bağımsız popülasyonun ortalamalarının karşılaştırılması
  - Şartlar: popülasyonlar normal yada normale yakın bir dağılım serğilemeli
  - Popülasyonlar birbirinden bağımsız olmalı, ve populasyonların standart sapmalari benzer olmalı
- Bir veri çiftinin karşılaştırılması

t dağılımlarında 2 birbirinden bağımsız popülasyonun ortalamalarının karşılaştırılması

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

n_1:1.ornekteki veri sayisi

s_p^2 = populasyon varyansinin birlestirilmis tahmini

s_1^2 = 1. ornegin varyansi

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 1}$$

df = serbeslik derecesi = $n_1 + n_2 - 2$

t dağılımlarında bir veri çiftinin karşılaştırlması

Örnek: bir fiziksel zayıflama kursuna katılmış kişilerinin kurs öncesi ve kursu sonrası kilolarının karşılaştırılması

Amaç: kursun etkinliğini test etmek

$$t = \frac{\overline{d}}{s_d / \sqrt{n}}$$

$df = n - 1$

$$s_d = \sqrt{\frac{\sum d^2 - \left(\frac{\left(\sum d\right)^2}{n}\right)}{n - 1}}$$

$\overline{d}$ = bir cift gozlemler arasindaki farkin ortalamasi

n: cift gozlem sayisi

s_d = bir cift gozlemlerin degerleri arasimdaki farkin standart sapmasi

Kategorisel veri analizi
ki-kare testi

Ki-kare testi (χ^2)

İki değişken arasında bir ilişki olup olmadığının testi

Örnek:

Kişiliğin deprosyonla bir ilişkisi var mıdır, yoksa bu iki değişken birbirinden bağımsızmıdır.

		Depresif mi?	
		evet	*hayır*
Kişilik	*dışadönük*	10	20
	içedönük	20	10

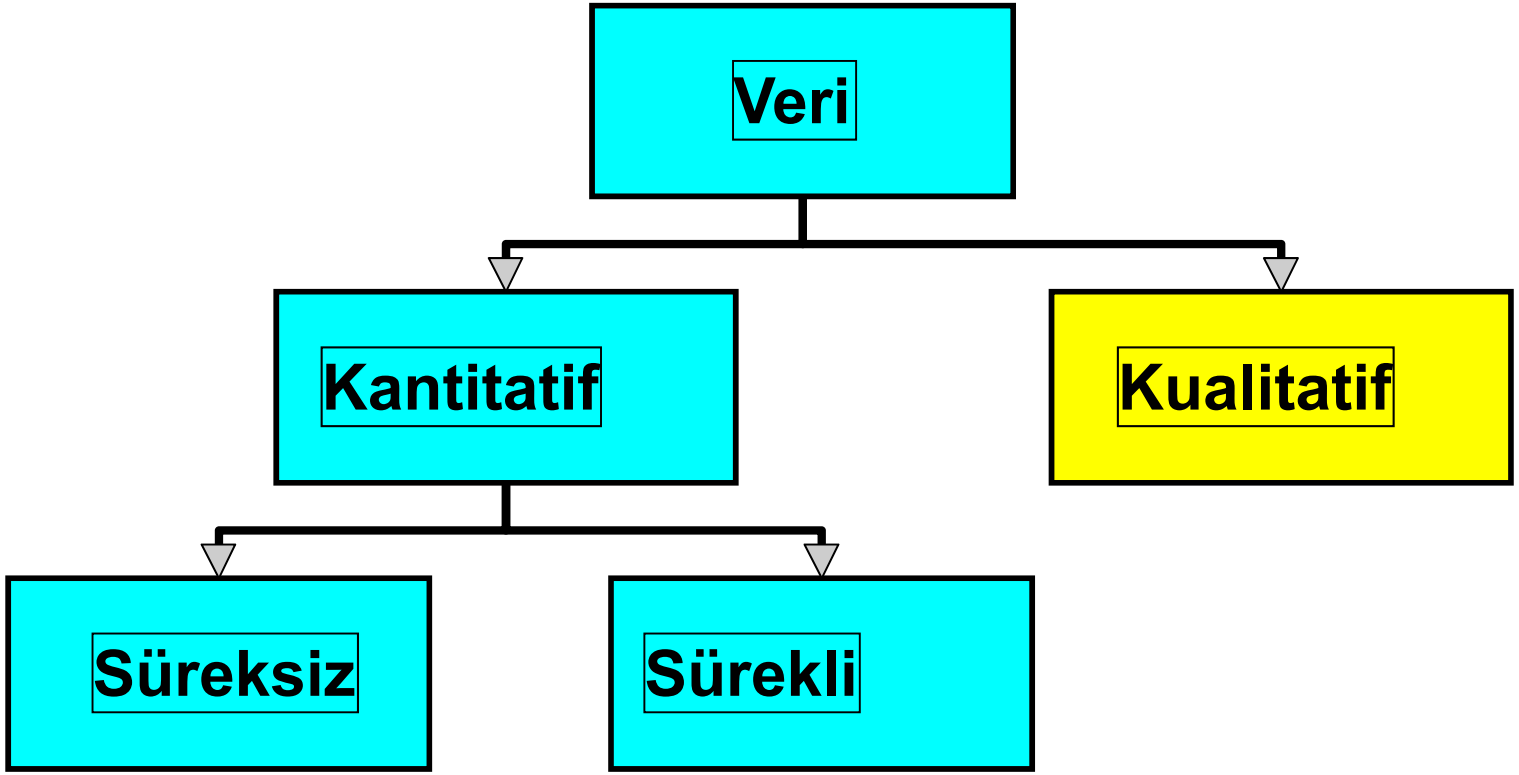
"hücreler"

<u>Not</u>:

→Değişkenler kategorik

→Her katılımcı sadece bir hücrede gözükür

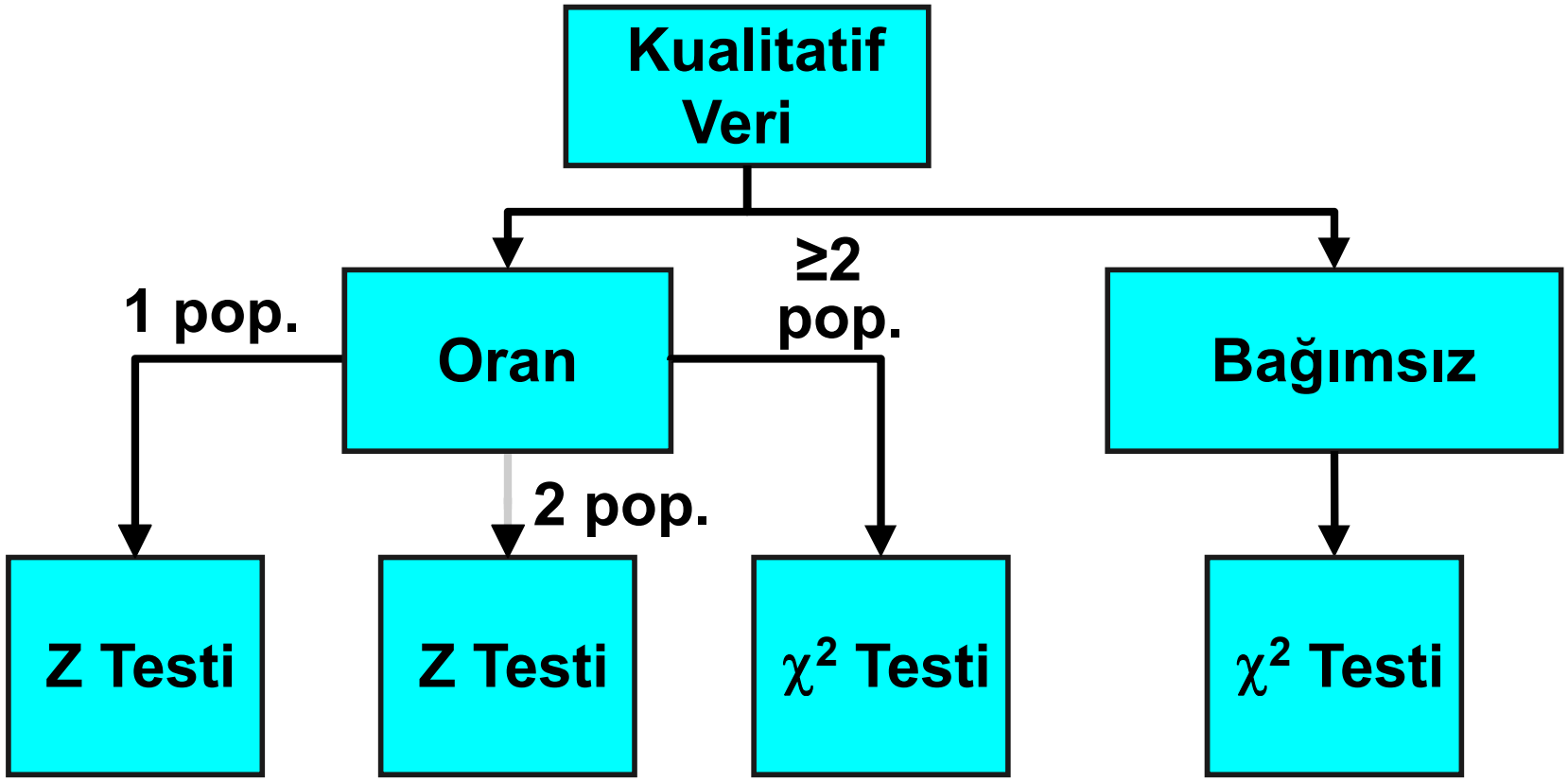
VERİ TÜRLERİ



Kualitatif Veri

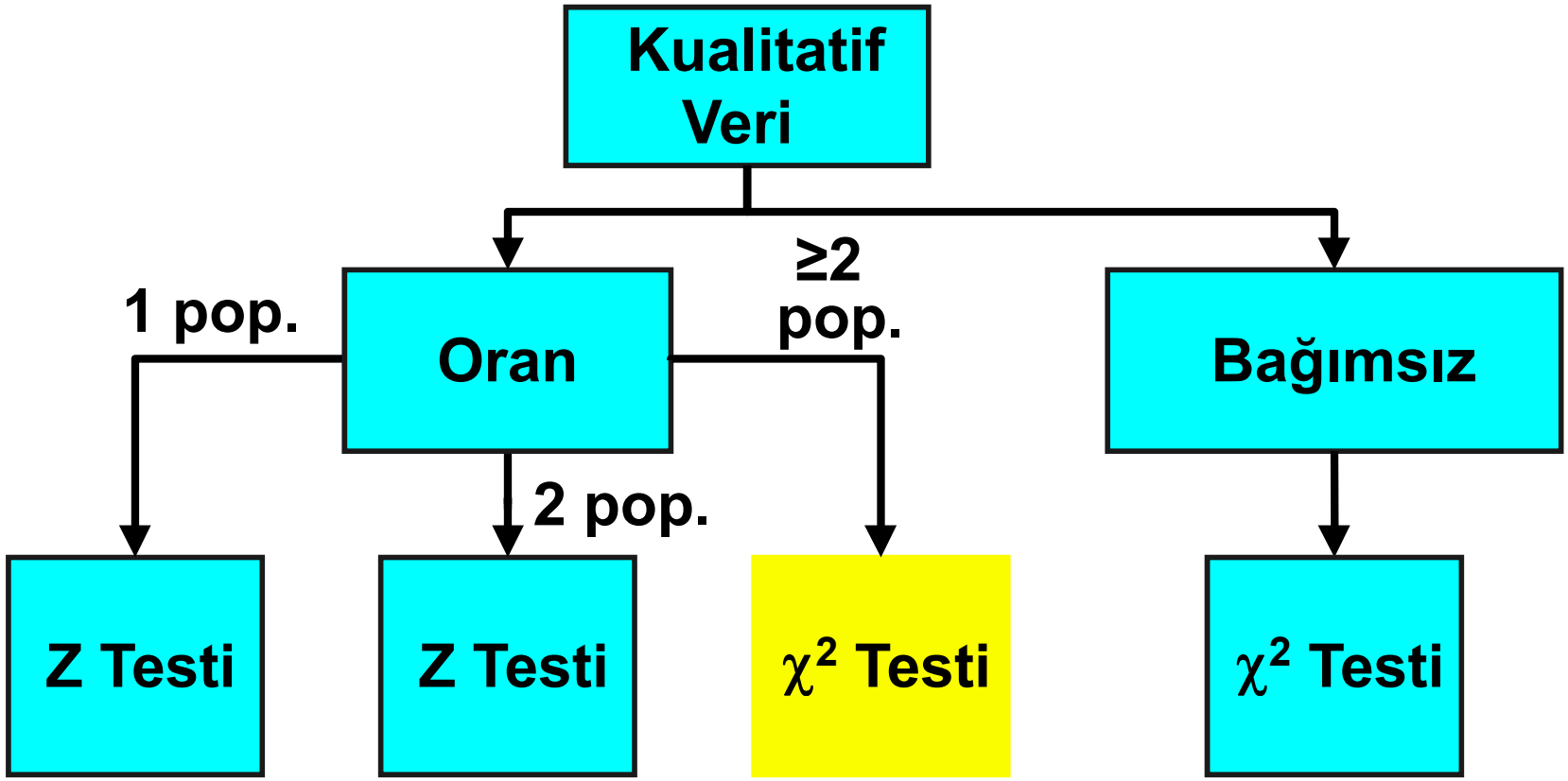
1. Kualitatif random değişkenler sınıflanabilen yanıtlar vermektedir.
   - Örnek: cinsiyet (Erkek, Kız)
2. Ölçüm kategorideki veri sayısını yansıtır
3. Nominal yada Ordinal ölçek

Hipotez testleri
Kualitatif Veri



Ki-kare (χ^2) Testi
k oranları için

Hipotez testleri
Kualitatif Veri



Ki-kare (χ^2) Testi
k Oranları için

1. Oranların sadece eşitliklerini test eder.
   - Örnek: $p_1 = 0.2$, $p_2=0.3$, $p_3 = 0.5$
2. Bir kaç seviyeli bir değişken
3. Varsayımlar
   - Multinomial Deney
     - Beklenen sayı ≥ 5
4. Bir yönlü olasılık tablasunu kullanmakta

Multinomial Deneyler

1. *n* sayıda benzer deneme
2. Her bir denemede k sayıda sonuç
3. Sabit sonuç oranları, p_k
4. Bağımsız denemeler
5. Random değişken sayıdır, n_k
6. Örnek; 100 (n) kişiye 3(k) adaydan hangisine oy vereceklerini sormak

Tek yönlü olasılık tablosu

1. k sayıda bağımsız grup içindeki (sonuçlar veya değişken seviyeleri) gözlem sayılarını gösterir

Tek yönlü olasılık tablosu

1. k sayıda bağımsız grup içindeki (sonuçlar veya değişken seviyeleri) gözlem sayılarını gösterir

Sonuçlar (*k* = 3)

ADAY			
Ali	**Veli**	**Aycan**	**Toplam**
35	**20**	**45**	**100**

Yanıt sayısı

χ^2 Testi,k sayıda oran için Hipotezler ve istatistik

1. Hipotezler
   - H_0: $p_1 = p_{1,0}$, $p_2 = p_{2,0}$, ..., $p_k = p_{k,0}$
   - H_a: p_i lar birbirine eşit değildir.

Hipotez edilen olasılık

χ^2 Testi,k sayıda oran için Hipotezler ve istatistik

1. Hipotezler
   - H_0: $p_1 = p_{1,0}$, $p_2 = p_{2,0}$, ..., $p_k = p_{k,0}$
   - H_a: p_i lar birbirine eşit değildiler

Hipotez edilen olasılık

2. Test istatistiği

Gözlemlenen sayı

$$\chi^2 = \sum \frac{[n_i - E[n_i]]^2}{E[n_i]}$$

Beklenen sayı

χ^2 Testi,k sayıda oran için Hipotezler ve istatistik

1. Hipotezler
   - H_0: $p_1 = p_{1,0}$, $p_2 = p_{2,0}$, ..., $p_k = p_{k,0}$
   - H_a: p_i lar birbirine eşit değildiler

Hipotez edilen olasılık

2. Test istatistiği

Gözlemlenen sayı

$$\chi^2 = \sum \frac{[n_i - E(n_i)]^2}{E(n_i)}$$

Beklenen sayı

3. Serbestlik derecesi, df= k-1

k, veri sayısı

χ^2 Testi Ana fikiri

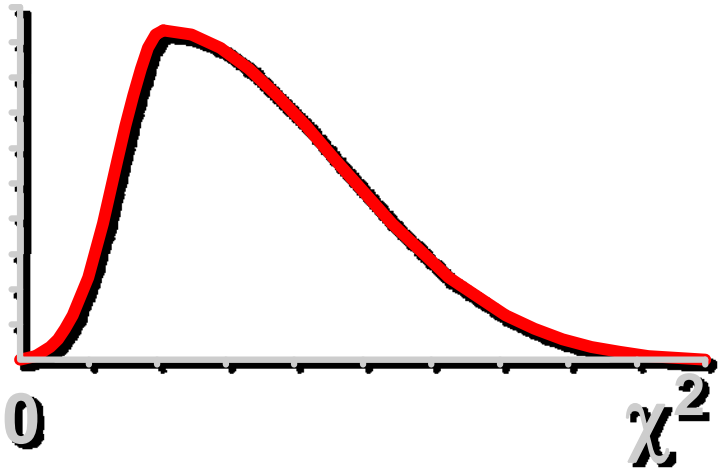
1. Null hipotezi doğru ise gözlemlenen sayıyla beklenen sayıyı karşılaştırır.
2. Gözlemlenen sayı nekadar beklenen sayıya yaklaşırsa null hipotezinin doğru olma olasılığı o kadar fazladır
   - Beklenen sayıyla arasındaki farkın karesi ile ölçülür.
     - Büyük değerler reddedilir.

Kritik değerin bulunması Örnek

k **= 3, ve** α **=0,05 için kritik** χ^2 **değeri nedir**

Kritik değerin bulunması Örnek

k **= 3, ve α =0,05 için kritik χ^2 değeri nedir?**



χ^2 **Tablo**

	Üst Kuyruk Alanı				
df	**0.995**	**…**	**0.95**	**…**	**0.05**
1	**...**	**…**	**0.004**	**…**	**3.841**
2	**0.010**	**…**	**0.103**	**…**	**5.991**

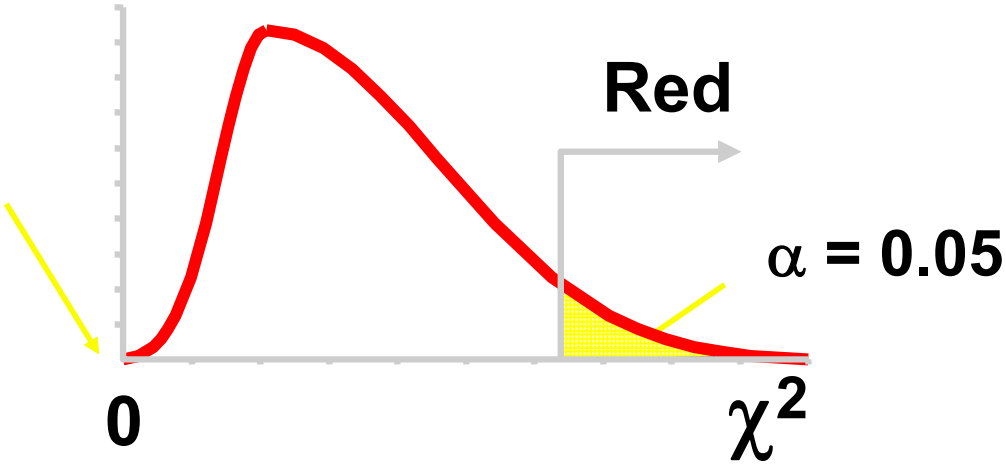
Kritik değerin bulunması Örnek

k **= 3, ve α =0,05 için kritik** x^2 **değeri nedir?**

Eğer $n_i = E(n_i)$,

$\chi^2 = 0$

H_0 **reddedilmez**



χ^2 **Tablo**

	Üst kuyruk alanı				
df	0.995	…	0.95	…	0.05
1	...	…	0.004	…	3.841
2	0.010	…	0.103	…	5.991

Finding Critical Value Example

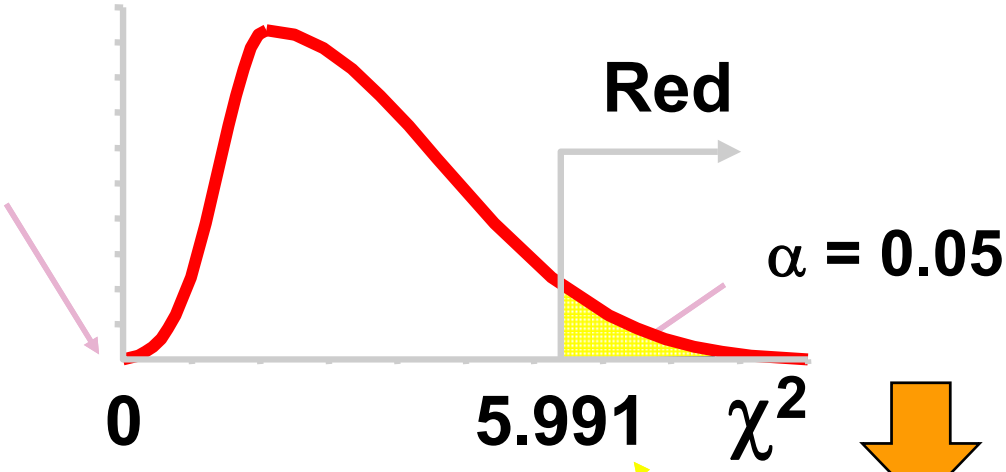
k **= 3, ve α =0,05 için kritik** x^2 **değeri nedir?**

Eğer $n_i = E(n_i)$,

$\chi^2 = 0$.

H_0 **reddedilmez**

df = $k - 1 = 2$



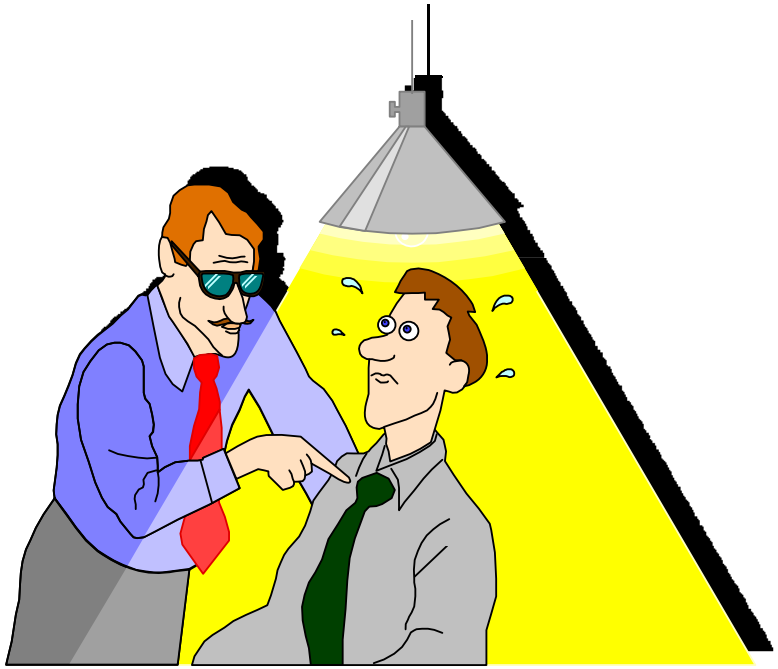
χ^2 **Tablo**

	Üst kuyruk alanı				
df	**0.995**	**…**	**0.95**	**…**	**0.05**
1	**...**	**…**	**0.004**	**…**	**3.841**
2	**0.010**	**…**	**0.103**	**…**	**5.991**

χ^2 Testi , k oranları için Örnek

İnsan kaynakları müdürü olarak, 3 farklı performans değerlendirme metodunun dürüstlük anlayışını test etmek istemetedir.

180 tane çalışan arasından, **63 ü 1. Methodu** dürüst olarak; **45 i 2. Methodu** dürüst olarak; **72** si ise **3. Methodu** dürüst olarak değerlendirmiştir. 0.05 risk derecesinde, çalışanların metodların dürüstlük derecesini algılamada bir farklılık varmıdır?



χ^2 Testi , *k* oranları için Örnek

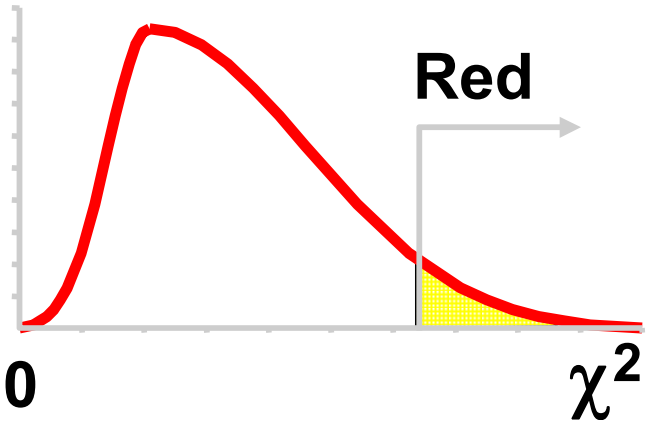
H_0:

H_a:

α =

n_1 = n_2 = n_3 =

Kritik değer(ler):



Test istattiği:

Karar:

Sonuç:

χ^2 Testi , k oranları için Örnek

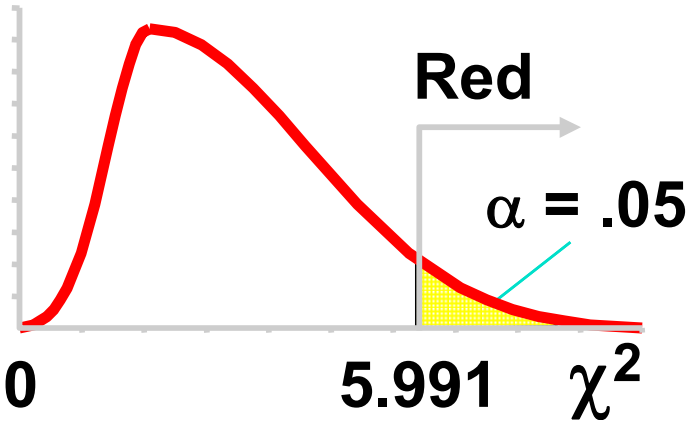
H_0: $p_1 = p_2 = p_3 = 1/3$

H_a: en az biri farklıdır

$\alpha = 0.05$

$n_1 = 63 \quad n_2 = 45 \quad n_3 = 72$

Kritik değer(ler):



Test istatistiği:

Karar:

Sonuç:

χ^2 Testi , k oranları için Çözüm

$$E[n_i] = np_{i,0}$$

$$E[n_1] = E[n_2] = E[n_3] = 180[1/3] = 60$$

$$\chi^2 = \sum \frac{[n_i - E[n_i]]^2}{E[n_i]}$$

$$= \frac{[n_1 - 60]^2}{60} + \frac{[n_2 - 60]^2}{60} + \frac{[n_3 - 60]^2}{60}$$

$$= \frac{[63 - 60]^2}{60} + \frac{[45 - 60]^2}{60} + \frac{[72 - 60]^2}{60} = 6.3$$

χ^2 Testi , *k* oranları için Çözüm

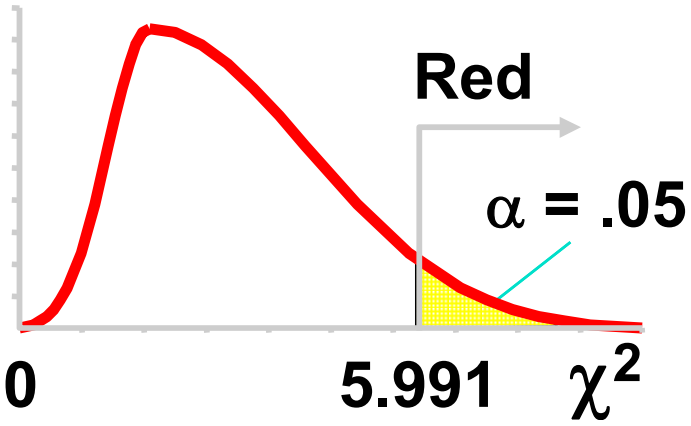
H_0: $p_1 = p_2 = p_3 = 1/3$

H_a: en az biri farklıdır

$\alpha = 0.05$

$n_1 = 63 \quad n_2 = 45 \quad n_3 = 72$

Kritik değer(ler):



Test istatistiği:

$\chi^2 = 6.3$

Karar:

Null hipotezi reddedilir, $\alpha = .05$

Sonuç:

Oranlar arasında bir farklılık olduğu delili mevcuttur.

χ^2 Testi, bağımsız

χ^2 Testi, bağımsız

1. İki kualitatif değişken arasında bir ilişkinin mevcut olup olmadığını gösterir
   - Bir örnek seçilir
   - Sebeb sonuç ilişkisi göstermez
2. Varsayımlar
   - Multinomial deney
   - tüm sayılar ≥ 5
3. Çift yönlü olasılık tablosu kullanır

χ^2 Test

Bir örnekten müşterek iki kualitatif değişkene ait gözlem sayısını gösterir

2. **değişken derecesi**

	Ev lokasyonu		
Ev stili	**Şehir**	**Kırsal**	**Toplam**
Apartman	**63**	**49**	**112**
Çiflik	**15**	**33**	**48**
Toplam	**78**	**82**	**160**

1. değiken derecesi

χ^2 Testi, bağımsız Hipotezler & İstatistik

1. Hipotezler
   - H_0: Değişkenler bağımsız
   - H_a: Değişkenler birbiriyle ilişkili (Bağımlı)
2. Test İstatistiği

Gözlenen sayı

$$\chi^2 = \sum \frac{\left[n_{ij} - \hat{E}(n_{ij})\right]^2}{\hat{E}(n_{ij})}$$

Beklenen sayı

Satırlar **Sütünlar**

Serbestlik derecesi: $(r - 1)(c - 1)$

χ^2 Testi, bağımsız beklenen sayılar

1. İstatiksel olarak bağımsız demek, birleşik olasığın marjinal olasılıklarının çarpımına eşit olduğu analamına gelmektedir.
2. Marjinal olasılıklar hesaplanır ve birleşik olasılık hesabı için çarpılır
3. Beklenen sayı= veri sayısı x birleşik olasılığa eşittir.

Beklenen sayıhesaplaması

$$\text{Beklenen sayı} = \frac{\text{Satır toplam} * \text{Sütün toplam}}{\text{Veri sayısı}}$$

χ^2 Testi, bağımsız Örnek

Birleşik olasılık = $\frac{116}{286}\frac{132}{286}$

$\frac{116}{286}$

	Diet Pepsi		
Diet Kola	Hayır Gözl.	Evet Gözl.	Toplam
Hayır	84	32	116
Evet	48	122	170
Toplam	132	154	286

$\frac{132}{286}$ Marjinal olasılık

Beklenen sayı $= \cancel{286} \cdot \frac{116}{\cancel{286}}\frac{132}{286}$

$=53.5$

χ^2 Testi, bağımsız Örnek

Pazarlama araştırması yapan bir analistsiniz.Rastgele seçtiğiniz 286 müşteri üzerinde yapacağınız araştırmada, müşterilere diet pepsi mi yada diet kola mı satın alıdıklarını soruyorsunuz.**α=0.05** risk derecesinde,ikisi arasında bir ilişki olduğuna dair delil varmıdır?

	Diet Pepsi		
Diet Kola	**Hayır**	**Evet**	**Toplam**
Hayır	**84**	**32**	**116**
Evet	**48**	**122**	**170**
Toplam	**132**	**154**	**286**

χ^2 Testi, bağımsız Çözüm

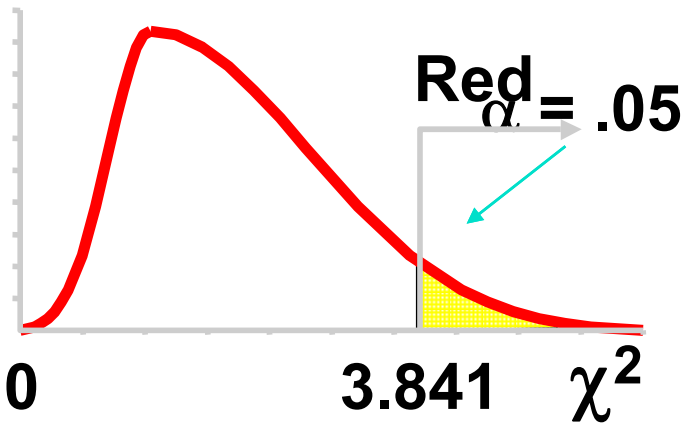
H_0: ilişki yok

H_a: ilişkili

α = 0.05

df = (2 - 1)(2 - 1) = 1

Kritik değer(ler):



χ^2 Testi, bağımsız Çözüm

✓ $E(n_{ij}) \geq 5$ tüm hücrelerde

$\frac{116 \cdot 132}{286}$ $\frac{154 \cdot 132}{286}$

	Diet Pepsi				
	Hayır		Evet		
Diet Kola	Gözl.	Bekl.	Gözl.	Bekl.	Toplam
Hayır	84	53.5	32	62.5	116
Evet	48	78.5	122	91.5	170
Toplam	132	132	154	154	286

$\frac{170 \cdot 132}{286}$ $\frac{170 \cdot 154}{286}$

χ^2 Testi, bağımsız Çözüm

$$\chi^2 = \sum \frac{\left[n_{ij} - \hat{E}[n_{ij}]\right]^2}{\hat{E}[n_{ij}]}$$

$$= \frac{\left[n_{11} - \hat{E}[n_{11}]\right]^2}{\hat{E}[n_{11}]} + \frac{\left[n_{12} - \hat{E}[n_{12}]\right]^2}{\hat{E}[n_{12}]} + \ldots + \frac{\left[n_{22} - \hat{E}[n_{22}]\right]^2}{\hat{E}[n_{22}]}$$

$$= \frac{[84 - 53.5]^2}{53.5} + \frac{[32 - 62.5]^2}{62.5} + \ldots + \frac{[122 - 91.5]^2}{91.5} = 54.29$$

χ^2 Testi, bağımsız Çözüm

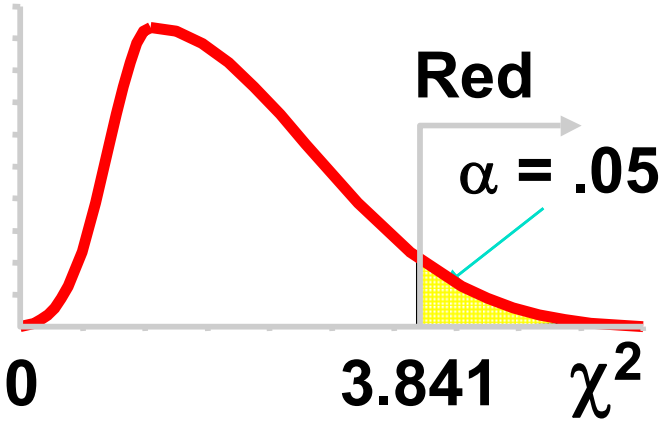
H_0: ilişki yok

H_a: ilişki var

$\alpha = 0.05$

df = (2 - 1)(2 - 1) = 1

Kritik değer(ler):



Test istatistiği:

$\chi^2 = 54.29$

Karar:

Null hipotezi reddedilir

α =0 .05

Sonuç:

Bir ilişki olduğuna

dair delil var

KORELASYON VE REGRESYON ANALİZİ

Dr. İrfan Yolcubal
Kocaeli Üniversitesi
Jeoloji Müh. Bölümü

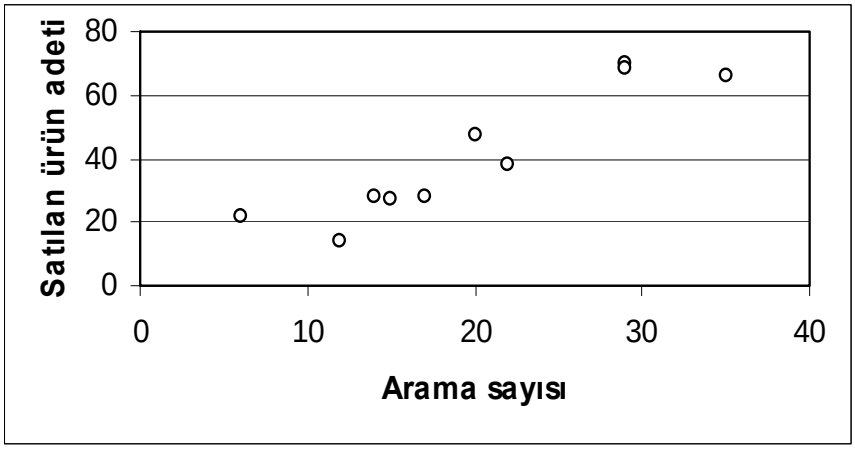
Korelasyon Analizi

- İki değişken arasındaki ilişkinin yada korelasyonunun derecesini belirlemek için kullanılan istatistiksel yöntem.
- **Bağımlı Değişken:** Tahmin edilen yada hesaplanan değişken
- **Bağımsız Değişken:** Tahmin için kullanılan değişken
- **Dağılım grafikleri:** 2 değişken arasındaki ilişkiyi gösteren grafikler

Korelasyon Analizi: Örnek

Satış elemanı	Telefonla yapılan arama sayısı	Satış yapılan ürün adeti
Ali	14	28
Veli	35	66
Ayşe	22	38
Gül	29	70
Hüsnü	6	22
Necati	15	27
Zehra	17	28
Fatma	20	47
Zeynep	12	14
Ahmet	29	68

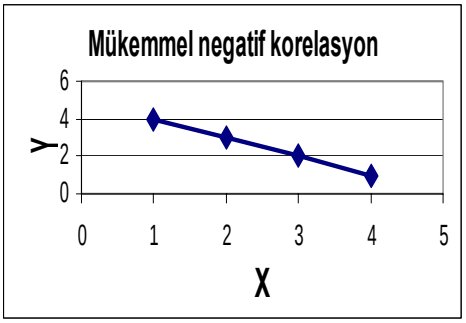
Dağılım Grafikleri



• Genellikle, bağımlı değişken: y ekseninde , bağımsız değişkende x ekseninde yeralır.

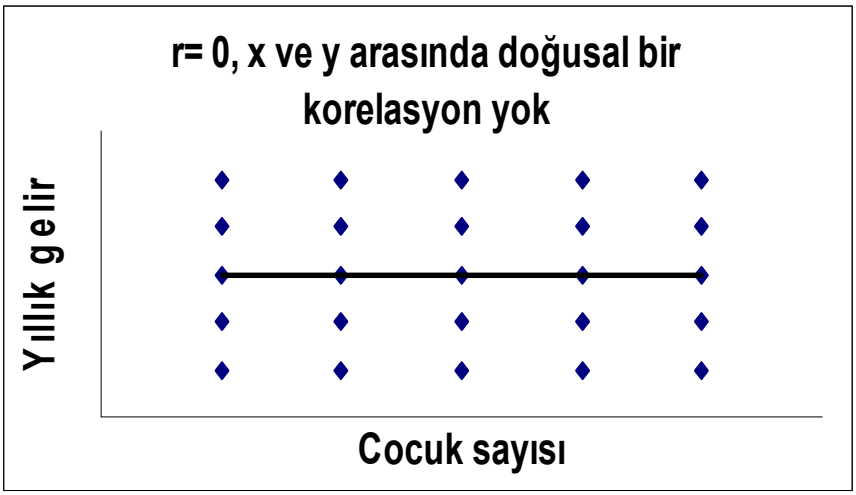
Korelasyon Katsayısı (r)

- 1900 yılında Karl Pearson tarafından tanımlandı.
- İki değişken arasındaki dogruasal korelasyonunun derecesini belirleyen bir katsayısı. $-1 \leq r \leq 1$
- r = -1 yada +1, iki değişken arasındaki korelasyonun mükemmel olduğunu ifade etmektedir.



r= -1 **r =1**

Korelasyon Katsayısı (r)



Korelasyon katsayısı (r)

Korelasyonun derecesi 2 değişken arasındaki ilişkinin yönüne(+ yada -) bağlı değildir.

Korelasyon katsayısının hesaplanması (r)

$$r = \frac{n\left(\sum XY\right) - \left(\sum X\right)\left(\sum Y\right)}{\sqrt{\left[n\left(\sum X^2\right) - \left(\sum X\right)^2\right]\left[\left(n\left(\sum Y^2\right) - \left(\sum Y\right)^2\right]}}$$

n : gozlem sayisi

Korelasyon Katsayısının Hesaplanması: Örnek

Satış elemanı	Telefonla yapılan arama sayısı(X)	Satış yapılan ürün adeti (Y)	X^2	XY	Y^2
Ali	14	28	196	392	784
Veli	35	66	1225	2310	4356
Ayşe	22	38	484	836	1444
Gül	29	70	841	2030	4900
Hüsnü	6	22	36	132	484
Necati	15	27	225	405	729
Zehra	17	28	289	476	784
Fatma	20	47	400	940	2209
Zeynep	12	14	144	168	196
Ahmet	29	68	841	1972	4624
TOPLAM	**199**	**408**	**4681**	**9661**	**20510**

$$r = \frac{10(9661) - (199)(408)}{\sqrt{\left[10(4681) - (199)^2\right]\left[10(20510) - (408)^2\right]}} = 0{,}924$$

Korelasyon Katsayısının Öneminin Test edilmesi

Küçük örneklemelerde (n<30)

Örnek: Önceki örnekte satılan ürün sayısı ile telofonla yapılan arama sayısı arasında güçlü bir korelasyonun olduğuna değindik. Fakat bu korelasyon sadece 10 örneğe dayanılarak çıkan bir sonuç. Gerçekte popülasyonun korelasyonu sıfır olabilir mi bilmiyoruz. Bu nedenle korelasyon katsayısının güvenirliliğini test etmemiz gerekmektedir.

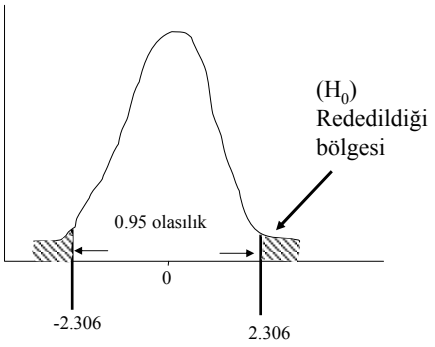
Hipotez: H_0: $\rho = 0$ H_1: $\rho \# 0$
ρ= popülasyonun korelasyonu

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.924\sqrt{10-2}}{\sqrt{1-(0.924)^2}} = 6.835$$

$$df = n - 2 = 10 - 2 = 8$$

% 5 risk ile

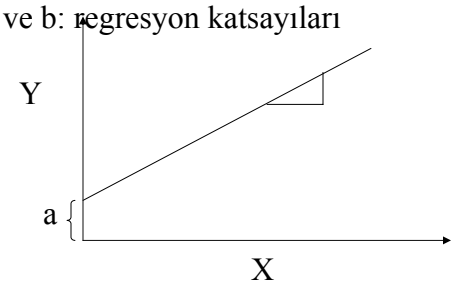
p<0.001



Regresyon Analizi

İki değişken arasındaki korelasyonun matematiksel ifadesini tespit etmek için yapılan analiz. Bu matematiksel ifadeye de regresyon denklemi diyoruz.

Regresyon denklemi genel ifadesi: Y` = a + bX
X: seçilen bağımsız değişkenin değeri
Y`: seçilmiş X değerine için tahmin edilen Y değeri
a: dogrunun Y ekseni kestiği noktanın değeri
b: doğrunun eğimi
a ve b: regresyon katsayıları

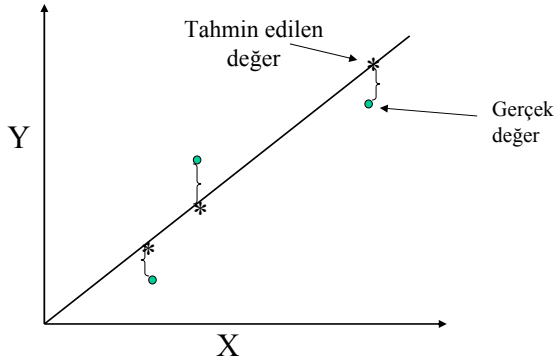


Regresyon Analizi

Regresyon denklemini tespit etmek için minimum kare (least square principle) prensibini kullanıyoruz.

Minumum Kare Prensibi: Gerçek Y değerleri ile tahmin edilen Y değerleri arasındaki düşey uzaklıkların karelerinin toplamını minumuma çekerek regresyon denklemini belirleme işlemi.

$$\sum (Y - Y`)^2 = \text{minimum}$$



Regresyon Katsayılarının (a ve b) Belirlenmesi

$$a = \frac{\sum Y}{n} - b\frac{\sum X}{n}$$

$$b = \frac{n\left(\sum XY\right) - \left(\sum X\right)\left(\sum Y\right)}{n\left(\sum X^2\right) - \left(\sum X\right)^2}$$

X: Bağımsız değişkenin değeri
Y: Bağımlı değişkenin değeri
n: Örnekteki veri sayısı

Örnek: Regresyon Katsayılarının (a ve b) Belirlenmesi

Satış elemanı	Telefonla yapılan arama sayısı(X)	Satış yapılan ürün adeti (Y)	X^2	XY	Y^2
Ali	14	28	196	392	784
Veli	35	66	1225	2310	4356
Ayşe	22	38	484	836	1444
Gül	29	70	841	2030	4900
Hüsnü	6	22	36	132	484
Necati	15	27	225	405	729
Zehra	17	28	289	476	784
Fatma	20	47	400	940	2209
Zeynep	12	14	144	168	196
Ahmet	29	68	841	1972	4624
TOPLAM	**199**	**408**	**4681**	**9661**	**20510**

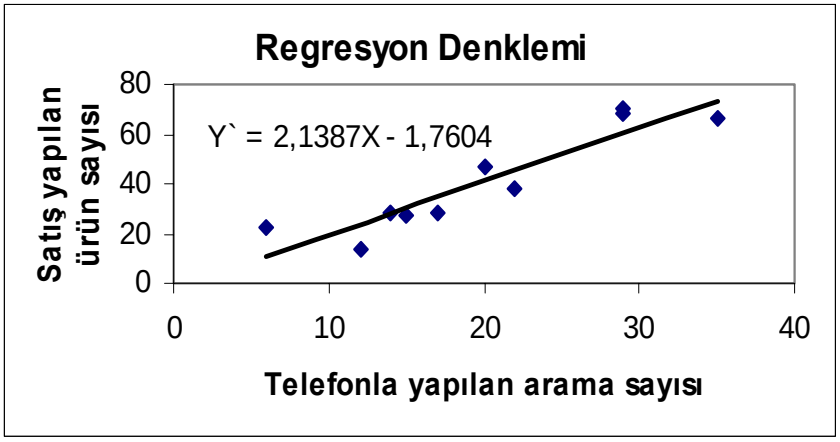
Örnek: Regresyon Katsayılarının (a ve b) Belirlenmesi

$$b = \frac{n\left(\sum XY\right) - \left(\sum X\right)\left(\sum Y\right)}{n\left(\sum X^2\right) - \left(\sum X\right)^2} = \frac{10(9661) - (199)(408)}{10(4681) - (199)^2} = 2{,}1387$$

$$a = \frac{\sum Y}{n} - b\frac{\sum X}{n} = \frac{408}{10} - 2{,}1387\ \frac{199}{10} = -1{,}706$$

$$Y` = -1{,}7601 + 2{,}1387$$

Örnek: Regresyon Doğrusunun grafiği



Regreson doğrusunun özelliği: $\sum (Y - Y`)^2 = \text{minimum}$

2.Regresyon doğrusu daima $\overline{X}$ $ve\ \overline{Y}$ değerlerinden geçmektedir.

Regresyon Katsayılarının Tahmininde Standart Hatanın($S_{y.x}$) Belirlenmesi

Standart Hata:Regresyon doğrusunun etrafında gözlenen değerlerin dağılımının yada yayılımınım ölçülmesi

$$S_{y.x} = \sqrt{\frac{\sum (Y - Y`)^2}{n-2}}$$

Sy.x= Tahminin standart hatası

y.x= x bağlı y değeri

$S_{y.x}$= 0 ise tüm noktalar regresyon doğrusu üzerine düşmektedir.

•Gözlem sayısının büyük olduğu durumlarda regresyon katsayılarının Tahmininde kullanılan daha pratik bir formül

$$S_{y.x} = \sqrt{\frac{\sum Y^2 - a\left(\sum Y\right) - B\left(\sum XY\right)}{n-2}}$$

Ornek: Regresyon Katsayılarının Tahmininde Standart Hatanın($S_{y.x}$) Belirlenmesi

Satış elemanı	Telofonla yapılan arama sayısı(X)	Satış yapılan ürün adeti (Y)	Tahmini satılan ürün adeti (Y`)	(Y-Y`)	$(Y-Y`)^2$
Ali	14	28	28,1817	-0,1817	0,033
Veli	35	66	73,0944	-7,0944	50,3305
Ayşe	22	38	45,2913	-7,2913	53,1631
Gül	29	70	60,2622	9,7378	94,8247
Hüsnü	6	22	11,0721	10,9279	119,4190
Necati	15	27	30,3204	-3,3204	11,0251
Zehra	17	28	34,5978	-6.5978	43,5310
Fatma	20	47	41,0139	-5,9861	35,8334
Zeynep	12	14	23,9043	-9,99043	98,0952
Ahmet	29	68	60,2622	7,7378	59,8735
TOPLAM	**199**	**408**		**0**	**566,1285**

$$Y` = -1,7601 + 2,1387 \qquad S_{y.x} = \sqrt{\frac{\sum (Y - Y`)^2}{n-2}} = \sqrt{\frac{566,1287}{10-2}} = 8.412$$

Doğrusal (linear) regresyon analiz uygulamak için varsayılan şartlar

1. Herbir X değeri için birden çokY değeri vardır. Bu Y değerleri normal dağılım göstermektedirler.
2. Bu Y değerlerinin ortalamaları daima doğrusal regresyon çizgileri üzerinde yeralır.
3. Bu normal dağılımların standart sapmaları birbirine eşittir.
4. Y değerleri istatiksel olarak birbirine bağlı değildir.

GÜVEN ARALIĞININ HESAPLANMASI

Güven aralığı tüm X değerlerine dayanan bir aralık olup seçilmiş bir X değeri için hesaplanan ortalama bir değerdir.

$$Y` \pm t(s_{y.x})\sqrt{\frac{1}{n}+\frac{\left(X-\overline{X}\right)^2}{\sum X^2-\frac{\left(\sum X\right)^2}{n}}}$$

X: seçilmiş X değeri
t: df=n-2 özgürlük derecisi için belli bir risk derecesine göre t dağılım tablolarından belirlenen t değeri

ÖRNEK: GÜVEN ARALIĞININ HESAPLANMASI

Örnek: 25 kez telefonla arama yapan satış uzmanlarının sattığı ürün sayısının güven aralığını hesaplayalım.Güven aralığını: %95 seçin

Satış elemanı	Telefonla yapılan arama sayısı(X)	Satış yapılan ürün adeti (Y)	X^2	XY	Y^2
Ali	14	28	196	392	784
Veli	35	66	1225	2310	4356
Ayşe	22	38	484	836	1444
Gül	29	70	841	2030	4900
Hüsnü	6	22	36	132	484
Necati	15	27	225	405	729
Zehra	17	28	289	476	784
Fatma	20	47	400	940	2209
Zeynep	12	14	144	168	196
Ahmet	29	68	841	1972	4624
TOPLAM	**199**	**408**	**4681**	**9661**	**20510**

ÖRNEK: GÜVEN ARALIĞININ HESAPLANMASI

$X = 25$

$Y` = 2{,}1387X - 1{,}7601$

$Y` = 51{,}7074$

%95 güven araligi

$\text{df} = \text{n} - 2 = 10 - 2 = 8$

$t = 2{,}306$(t dagilim tablolarindan)

$s_{y.x} = 8{,}412$

$$Y` \pm t(s_{y.x})\sqrt{\frac{1}{n} + \frac{(X-\overline{X})^2}{\sum X^2 - \frac{(\sum X)^2}{n}}} = 51{,}7074 \pm 2{,}306(8{,}412)\sqrt{\frac{1}{10} + \frac{(25-19{,}9)^2}{4681 - \frac{(199)^2}{10}}}$$

$51{,}7074 \pm 7{,}1558$

TAHMİN ARALIĞININ HESAPLANMASI

Tahmim aralığı belli bir X değeri için olup o X değerine karşılık gelen değerlerin aralığını verir.

$$Y`\pm t(s_{y.x})\sqrt{1+\frac{1}{n}+\frac{\left(X-\overline{X}\right)^2}{\sum X^2-\frac{\left(\sum X\right)^2}{n}}}$$

Örnek: 25 kez arama yapan Zekinin %95 güven aralıkla satış yapacağı ürün adetinin tahmini aralığının belirlenmesi