Örneklem Dağılımları ve Merkezi Limit Teoremi

Çıkarımsal İstatistik (Inferential Statistics) : Örneklemden yola çıkarak ana kütleyle (popülasyonla) ilgili çıkarımlarda bulunmak (Smidt, 2001)

İstatistiksel çıkarım; sınırlı sayıda birey, denek ya da nesneden elde edilen verilerden, bu verilerin alındığı evren hakkında bilgi edinmek (Hayram, 2011).

Çıkarımsal istatistik, örnekleme ve örnekleme dağılımı kavramı temeline dayanır. Örneklemenin anlamını anlamak, belki de istatistiği anlamakla eşdeğerdir. Tamsayım çoğu zaman imkânsızdır, pahalıdır ve çok zaman gerektirir. Örnekleme dağılımı kavramı istatistiğin en güç anlaşılan kavramlarından biridir. Örnekleme dağılımı kavramının iyi bir şekilde anlaşılmaması, bu kavram üstüne kurulan parametrelere ilişkin çıkarımların ve hipotez testlerinin anlaşılmamasına yol açma riski taşır (Gürsakal, 2013).

Normal dağılım neden önemlidir?

1) Hayvan ve bitkilerin birçok fiziksel özelliklerinin, eğitimde IQ testi gibi test sonuçlarının, kestirim ve ölçüm hatalarının, hisse senetlerinin getirileri gibi birçok veri normal dağılıma uymaktadır.
2) Bir veri kümesinin dağılımı normal olmasa bile bu dağılımdan elde edilecek örneklemlerin ortalamalarının (veya oranlarının) dağılımları normal dağılıma yaklaşmaktadır (Merkezi Limit Teoremi) (Gürsakal, 2013).

Örnekleme Dağılımı

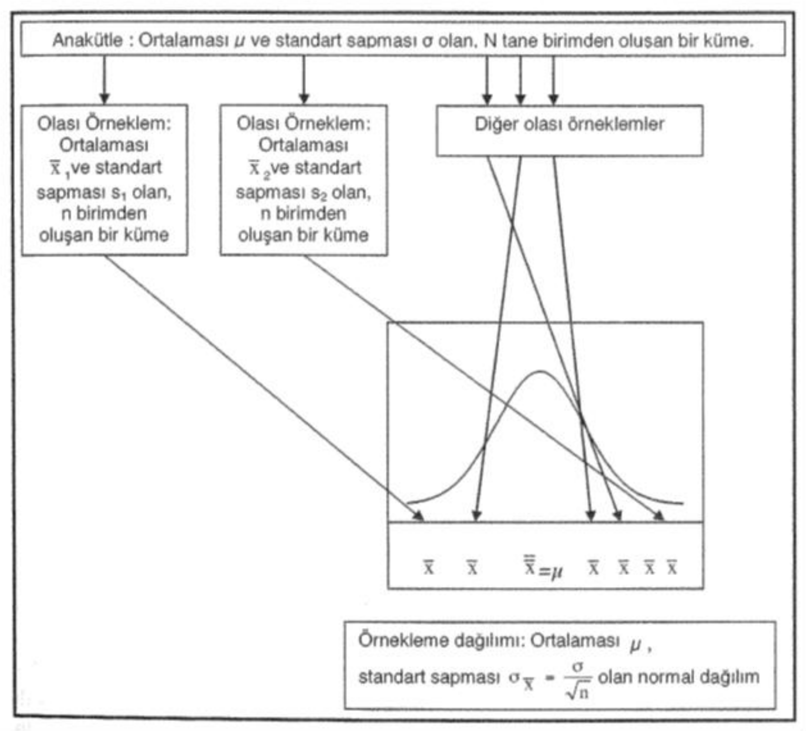
İstatistik dağılımları, istatistiğin değerinin hesaplanacağı örneğin alınmış olduğu ana kütlenin dağılımına bağlıdır. $X_1, X_2, X_3,\ldots X_n$ bir değişken vektörü olarak örneklemi ifade eder. Bu örneklemdeki her bir değişkenin dağılımı ana kütlenin dağılımı ile aynıdır. Bu örneklemin gerçekleşmiş hali $x_1, x_2, x_3,\ldots x_n$ örneğidir. Örneklemdeki değişkenlerin bir fonksiyonu olan değişken bir istatistik' tir (Armutlulu, 2008).

Bir istatistiğin değeri, anakütleden çekilecek belirli bir örnekleme bağlıdır. Bu nedenle, örneklem istatistiği bir rassal değişkendir ve bir olasılık dağılımına sahiptir. İstatistiğin olasılık dağılımına örneklem dağılımı (sampling distribution) adı verilir (Gürsakal, 2013).

Örnekleme dağılımları bir ana kütleden çekilebilecek tüm olası örneklemlerden yararlanılarak oluşturulan teorik dağılımlardır. Bu dağılımların da diğer dağılımlar gibi ortalaması ve standart sapması söz konusudur (Gürsakal, 2013).

-Ortalamaların örnekleme dağılımı (Sampling distribution of the mean)

-Oranların örnekleme dağılımları (sampling distribution of the proportion)



Şekil 1: Örnekleme Dağılımı Kavramı (Gürsakal, 2013)

Anakütle	Örneklem	Örnekleme dağılımı
μ : Anakütle ortalaması	$\bar{x}$: Örneklem ortalaması	$\bar{\bar{x}}$:Örnekleme dağılımının ortalaması
σ:Anakütle standart sapması	s: Örneklem standart sapması	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ Örnekleme dağılımının standart sapması
N: Anakütle büyüklüğü	n : Örneklem büyüklüğü	
π : Anakütle oranı	p : Örneklem oranı	

Tablo 1: Anakütle, örnekleme ve örnekleme dağılımı için notasyon (Gürsakal, 2013)

Örnek Ortalamasının Örnekleme Dağılımı

Ana kütlenin merkezi eğilimini belirlemede en güçlü istatistiklerden biri aritmetik ortalamadır. Ortalaması μ, varyansı σ^2 olup, toplam eleman sayısı N olan sonlu bir ana kütleden n birimlik örneklem tasarlandığında

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

İstatistiğinin örneklem dağılımının ortalaması, anakütle ortalamasına eşit, μ olup varyansı,

$$Var(\bar{X}) = \frac{\sigma^2}{n}\frac{N-n}{N-1}$$

Olur. Buradaki N-n / N-1 kesri sonlu ana kütle çarpanıdır. (Eğer anakütle örneğin en az 20 katı kadar büyükse sonlu anakütle çarpanı yaklaşık olarak bire eşittir, ancak anakütle büyüklüğü "N" örnek büyüklüğünün 20 katından az ise sonlu anakütle çarpanı 1' den küçük çıkar ve örneklem ortalamasının dağılımının standart sapması hesaplanırken bu çarpanı kullanmamaız gerekir (Orhunbilge, 2013)). Eğer ana kütle eleman sayısı bilinmeyip sonsuz kabul edilirse bu çarpan kullanılmaz. Bu durumda; X istatistiğinin dağılımının ortalaması μ, varyansı σ^2/n olur. Dağılımın şeklinin yol gösterici teoremlerinin başında Merkezi Limit Teoremi (MLT) gelir (Armutlulu, 2008)

Başka bir ifade ile anakütleden tesadüfi olarak n birimlik örnekler seçildiğinde, örnek ortalamalarının örnekleme dağılımı; ana kütleden çekilebilecek aynı büyüklükteki tüm mümkün örneklerin hepsinin örnek ortalamasının olasılık dağılımıdır (Orhunbilge, 2003).

Örnek; (Orhunbilge, 2013, Syf. 276)

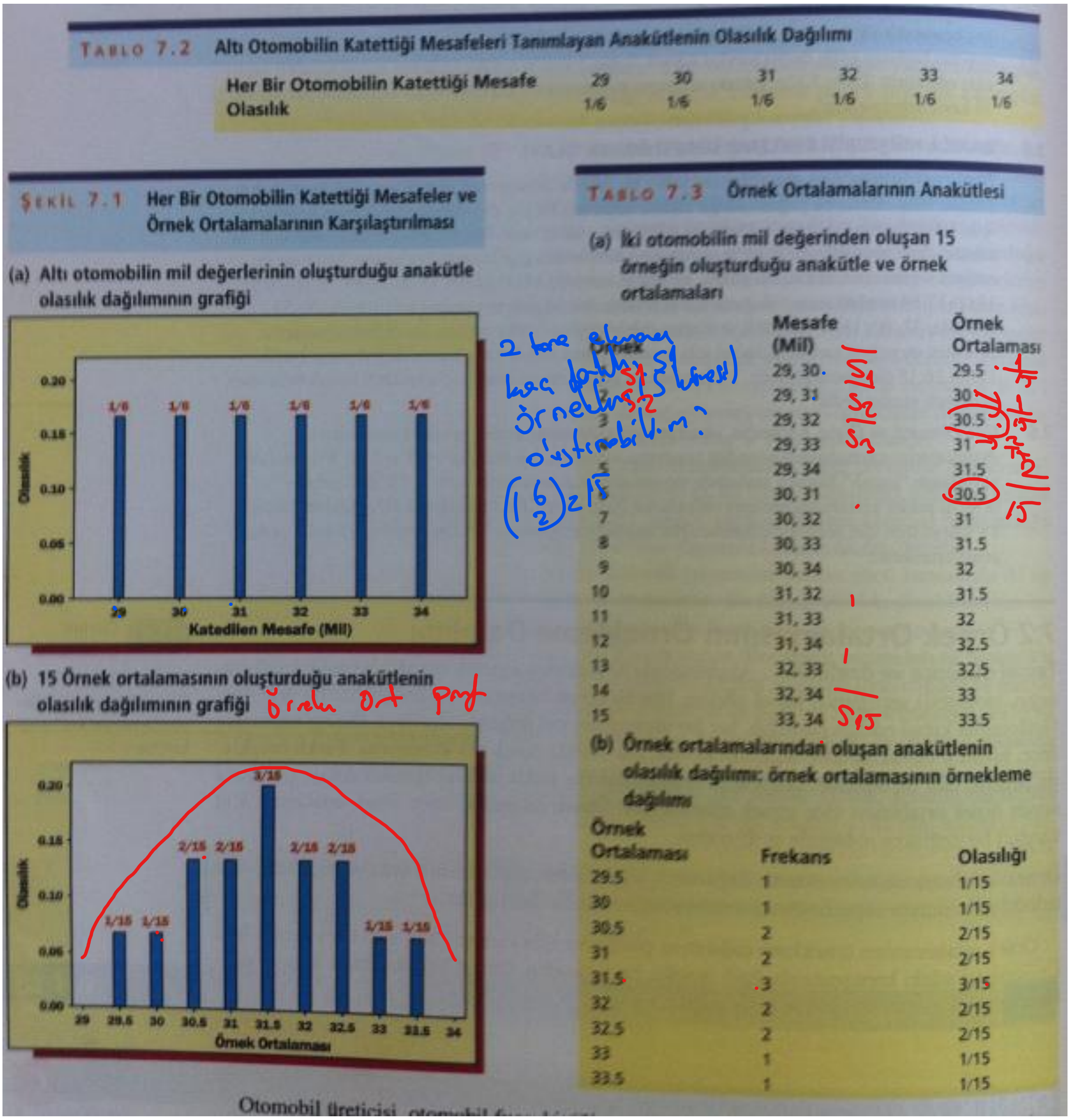
Bir otomobil firması, ortasınıf yeni bir modeli satışa sunmuştur. Yeni modelden altı tane deneme üretimi yapılmış ve EPA mil testi yapılmıştır. Test sonucunda otomobillerin mil değerleri 29, 30, 31, 32, 33 ve 34 olarak belirlenmiştir.

6 otomobillik ana kütleden tesadüfi olarak 2 otomobil seçildiğinde 15 örneklem elde edilecektir.

$\binom{6}{2} = \frac{6.5.4}{4.2.1} = 15$

Bu anakütlenin örneklem dağılımının grafiği aşağıda verilmiştir (Şekil 2).

Burada 6 araçtan oluşan anakütle küçük bir anakütledir. Bununla birlikte çoğu zaman incelemenin yapılacağı anakütle ya sonlu çok büyük ya da sonsuz anakütledir. Bu durumda bu anakütlenin ölçülmesi pratik olmayacak ya da imkansız olacaktır. Bunun yerine anakütleden tesadüfi olarak örnek seçer ve bu örneğin ortalamasını ana kütle ortalamasının tahmini olarak kabul kullanırız (Şekil 3).



TABLO 7.2 Altı Otomobilin Katettiği Mesafeleri Tanımlayan Anakütlenin Olasılık Dağılımı

Her Bir Otomobilin Katettiği Mesafe	29	30	31	32	33	34
Olasılık	1/6	1/6	1/6	1/6	1/6	1/6

ŞEKİL 7.1 Her Bir Otomobilin Katettiği Mesafeler ve Örnek Ortalamalarının Karşılaştırılması

(a) Altı otomobilin mil değerlerinin oluşturduğu anakütle olasılık dağılımının grafiği

(b) 15 Örnek ortalamasının oluşturduğu anakütlenin olasılık dağılımının grafiği

TABLO 7.3 Örnek Ortalamalarının Anakütlesi

(a) İki otomobilin mil değerinden oluşan 15 örneğin oluşturduğu anakütle ve örnek ortalamaları

Örnek	Mesafe (Mil)	Örnek Ortalaması
1	29, 30	29.5
2	29, 31	30
3	29, 32	30.5
4	29, 33	31
5	29, 34	31.5
6	30, 31	30.5
7	30, 32	31
8	30, 33	31.5
9	30, 34	32
10	31, 32	31.5
11	31, 33	32
12	31, 34	32.5
13	32, 33	32.5
14	32, 34	33
15	33, 34	33.5

(b) Örnek ortalamalarından oluşan anakütlenin olasılık dağılımı: örnek ortalamasının örnekleme dağılımı

Örnek Ortalaması	Frekans	Olasılığı
29.5	1	1/15
30	1	1/15
30.5	2	2/15
31	2	2/15
31.5	3	3/15
32	2	2/15
32.5	2	2/15
33	1	1/15
33.5	1	1/15

2 tane alırsam kaç farklı (15 farklı) örneklem oluşturabilirim?

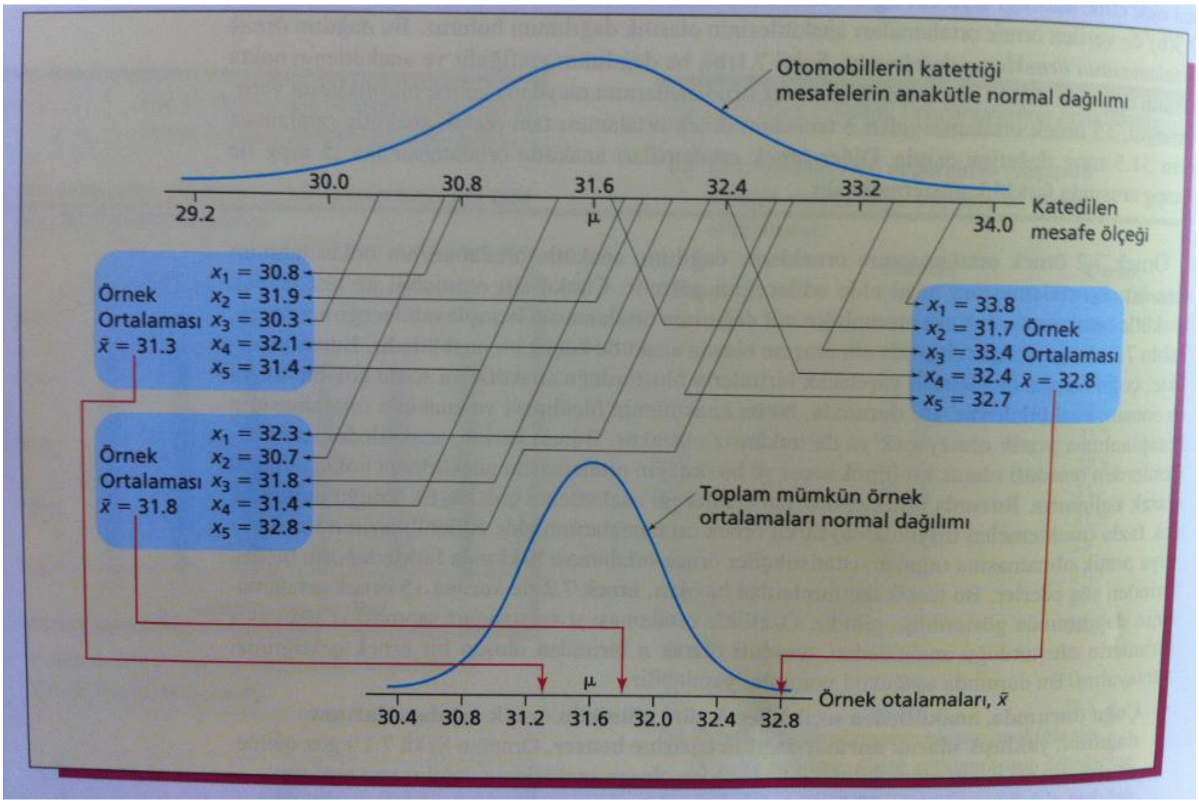
$\binom{6}{2} = 15$

örnek ort. prof

S_1 S_2 S_3 … S_{15}

Otomobil üreticisi ...

Şekil 2: Örnek (Orhunbilge, 2013)

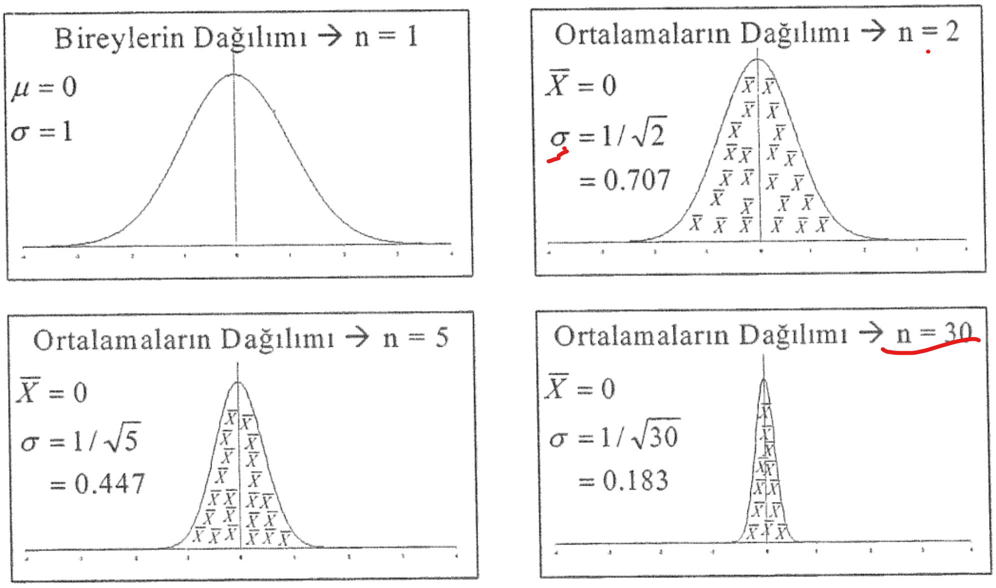


Şekil 3: Otomobillerin katettiği mesafelerin normal dağılımı ve mümkün bütün örnek ortalamalarının normal dağılımı (Orhunbilge, 2013)

Örnek 1 : Bu yıl üretilecek bütün yeni orta boy otomobillerin mesafe değerleri ana kütlesinin sonsuz bir ana kütle olduğunu düşünelim. Bu anakütlenin μ= 31 mil ve σ=.8 ile normal dağıldığını varsayarsak ve tesadüfi olarak n=50 büyüklüğünde bir örnek çektiğimizde ve EPA testi uyguladığımızda; örneğin ortalamasının 31,56 değerinden büyük olması olasılığı nedir? **(Orhunbilge, 2013)**

Merkezi Limit Teoremi

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



Şekil 2: Merkezi Limit Teoremi (6 Sigma Eğitim Notları, Matris Danışmanlık)

- MLT matematik olarak aşağıdaki denklemlere indirgenebilir:

$$\mu_{\text{Örneklerin Ortalamaları}} = \mu_{\text{AnaKütle}}$$

$$\mu_{\bar{x}} = \mu_x$$

Eşit örnek büyüklükleri varsayılarak $n_1=n_2=n_3$ vb.

$$\sigma_{\text{Örneklerin Ortalamaları}} = \frac{\sigma_{\text{Ana Kütle}}}{\sqrt{n}} \quad \text{yani} \quad \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{\sigma}{\sqrt{n}}$$

$$P(\text{Örneklerin Ortalamaları}) \approx \text{Normal} \quad (n \text{ büyüdükçe})$$

(Örnek büyüklüğü arttıkça, örnek ortalamalarının dağılımının normalliği artar)

Şekil 3: MLT (Matematiksel ifade) (6 Sigma Eğitim Notları, Matris Danışmanlık)

Merkezi limit teoremi normal dağılımların iki önemli özelliği ile ilgilidir.

1) İlgili X değişkenine göre yığının bölünümü normal ise örnek çapı ne olursa olsun örnekleme dağılımı da normaldir. Hatta n=1 iken örnekleme bölünümü ile yığına ilişkin bölünüm aynıdır. Örneklem çapı olan n arttıkça bölünüm tepe noktası sivrileşir.
2) İlgili X değişkenine göre yığının bölünümü normal değilse örnek çapı arttıkça örnekleme dağılımı da normale yaklaşır. Yani n'nin artması ile örnekleme bölünümünün normale yaklaşması arsında doğru orantı vardır. Örnek çapının hangi değerleri için örnekleme bölümünümünün normale yaklaşacağı yığına ilişkin bölünümün yatıklık derecesine bağlıdır. Aşırı yatık bölünümler için örnek çapının 30 dolayında olması örnek bölünümünün normalliği için yeterli sayılmaktadır (Çil, 2008)

Merkezi limit teoreminin geçerliliği sürekli rassal değişkenlerin toplamlarıyla sınırlı değildir. Kesikli rassal değişkenlere de genişletilebilir. (Şenesen, 2009)

Merkezi limit teoreminin yorumunda dikkatli olunmalıdır. Çünkü örneklem hacmi sonsuza giderken " σ^2 / √n " ifadesi sıfıra gitmektedir. Buradaki sonsuzluk yeterince büyük anlamındadır. Uygulamada anakütle ne şekilde dağılmış olursa olsun, n≥30 olduğunda x bar 'ın dağılımının normal dağılıma yakınsadığı görülmüştür (Armutlulu, 2008).

Anakütle ortalaması ile ilgili sonuç çıkarımında anakütlenin dağılımı ancak 30' dan küçük örneklerle çalışırken gündeme gelir. Örnek hacmi 30' dan küçükse ve ana kütlenin dağılımı normal varsayılırsa, varyansın bilinmesi durumunda normal dağılım; varyansın bilinmemesi durumunda t-dağılımı kullanılır. Örneklem hacmi 30' dan küçük olup anakütle dağılımı için normallik varsayımı yapılamıyorsa parametrik olmayan çalışmalar gündeme gelir (Armutlulu, 2008)

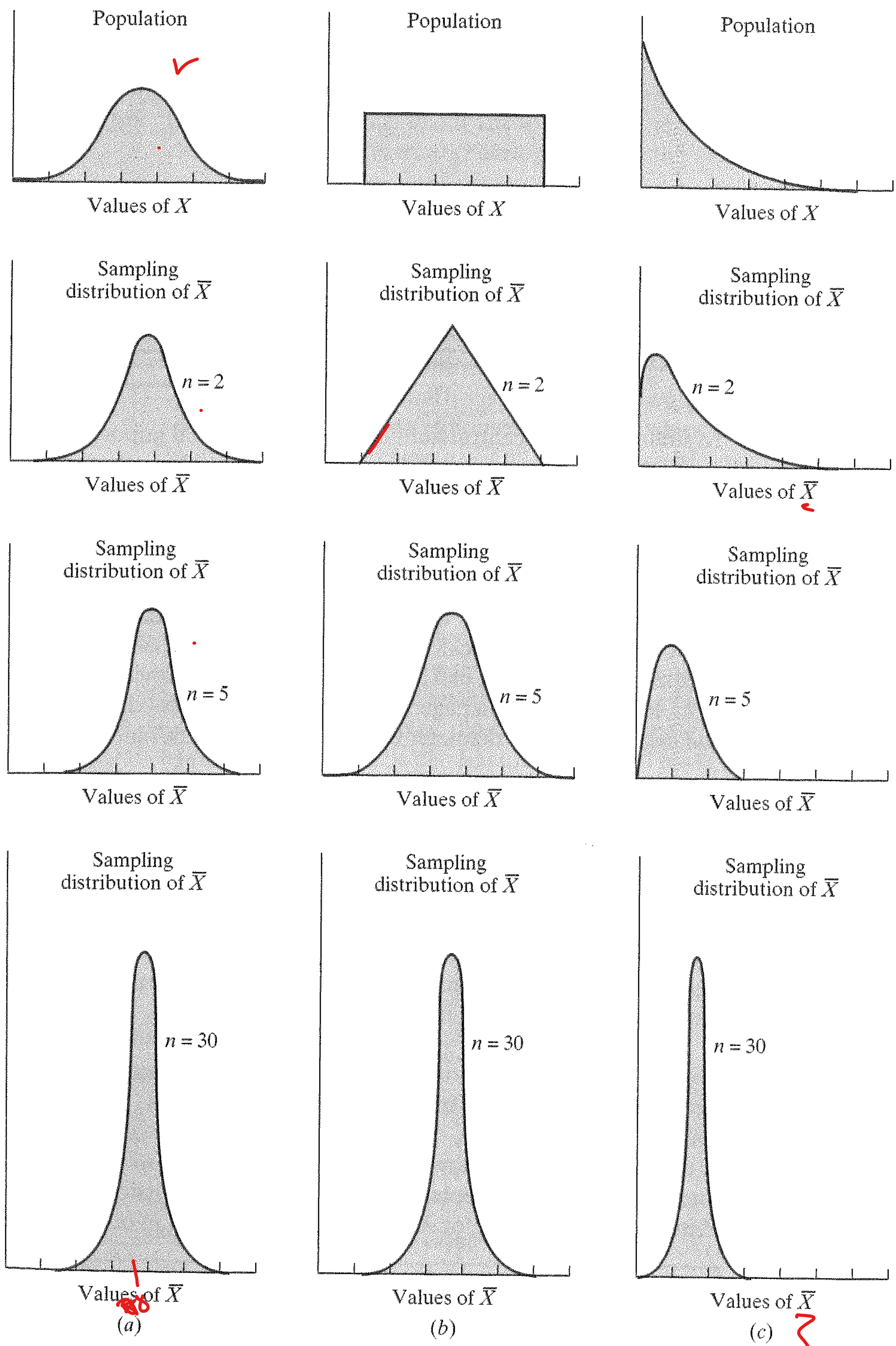
Sonuç olarak;

Anakütlenin dağılımına bakılmaksızın, n≥ 30 için örneklem ortalaması yaklaşık olarak normal dağılmaktadır. Eğer anakütle dağılımı simetriğe yakınsa, n ≥ 15 için örneklem ortalaması yaklaşık olarak normal dağılmaktadır. Eğer anakütle normal dağılıma sahipse, örneklem büyüklüğünden bağımsız olarak örneklem ortalaması normal dağılmaktadır (Smidt, 2001).

Örnek 3: Anakütle ortalaması 0,503 inch, standart sapmasının 0,004 inch olduğu biliniyorsa; bir ürünün çapının 0,503 ile 0,507 inch arasında olması olasılığı nedir?

Eğer popülasyondan n=4 büyüklüğünde bir örnek seçilirse, örneklemin ortalamasının 0,503 ile 0,507 inch arasında olması olasılığı nedir? (Smidt, 2001)

Örnek 4: Bir kamyon firmasının kullanmakta olduğu eski faturalama sisteminin ortalama ödeme süresi 39 güne eşittir. Danışmanlık firması yeni kuracağı sistemin ortalamasının 19,5 günden daha az olacağını iddia etmektedir. Yeni sistemi denemek için faturalar arasında tesadüfi olarak 65 fatura seçilmiştir. Bu 65 faturanın ödeme süresinin ortalaması 18,1077 gündür. Eğer ana kütlenin ödeme süresi ortalaması 19,5 günden daha az ise, ölçüm yapılan örneğin ödeme süresi ortalamasının 18,1077 güne eşit veya küçük olmasının olasılığı nedir? (σ= 4,2) (Orhunbilge, 2013)



Şekil 4: Farklı dağılımlara sahip ana kütlelerin örneklem dağılımları (smidt, 2001)

Tesadüfi olarak seçilen bir örneklemin ortalaması, ana kütle ortalamasından çok uzak bir değer olabilir. Genelde örneklemin ortalamalarının belli değerler arasında olup olmadığını incelemek yerine, ortalamaların belli oranının hangi değerler arasında olması gerektiği ile ilgileniriz (Smidt, 2001).

Örnek: 5 Anakütle ortalaması 0,503 inch, standart sapmasının 0,004 inch olduğu biliniyorsa, örnek ortalamalarının % 95'i n= 25 için hangi değerler arasındadır? (Smidt, 2001).

Oranlarla ilgili Örnekleme Dağılımı

Kategorik veri söz konusu olduğunda; başka bir ifadeyle bir özelliğin beklenen veya beklenmeyen durumları (başarı = 1, başarısızlık = 0) söz konusu olduğunda (Smidt, 2001) anakütlede istenen durumun oranı ile ilgili sonuç çıkarımında örneklemdeki oranın örnekleme dağılımı gündeme gelir. Deney veya gözlem sonuçları istenen-istenmeyen gibi iki kategoriden birine giriyorsa ardışık örnekleme bir Bernoulli süreci olacaktır. Ardışık Bernoulli denemelerinde istenen durumun gerçekleşme sayısının dağılımı da Binom dağılımı olacağından sonuç çıkarımı Binom dağılımından yapılır. Örneklem oranı için geliştirilen örnekleme dağılımının ortalaması p, standart hatası,

$$\sigma_{\bar{p}} = \sqrt{\frac{pq}{n}}$$

Eğer sonlu ana kütleden örneklem tasarlanmış ise ana kütle eleman sayısı N, örneklem hacmi n olmak üzere standart hata,

$$\sigma_{\bar{p}} = \sqrt{\frac{pq}{n}}\sqrt{\frac{N-n}{N-1}}$$

Örnekten hesaplanan oran p, anakütle oranı p olmak üzere np>5 ve n(1-p)>5 olursa p istatistiğinin dağılımı normal dağılıma yaklaşır ve bu durumda normal dağılım kullanarak çıkarımda bulunulur (Armutlulu, 2008).

Burada;

q = 1-p olduğuna dikkat edelim,

bazı kaynaklarda anakütle ortalaması π , örneklem ortalaması p ile gösterilirken

bazı kaynaklarda ana kütle ortalaması p ile örneklem ortalaması ise $\bar{p}$ veya $\hat{p}$ ile gösterilebilir.

Örnek 6: Bir gıda firması "kolay akıtılan" bir ağızlığa sahip plastik ambalaj içinde krem peynir pazarlamaktadır. Bu ağızlığın üretimi pahalıdır. Firma daha az pahalı bir ağızlık geliştirmiştir. Firma, yeni ağızlığın bazı müşterileri üründen soğutabileceğini düşünmektedir. Firma çalışması, yeni ürünün piyasaya girişi ile mevcut müşteriler içinden kaybedilebilecek müşterilerin oranı % 10' dan küçük olduğu zaman karın artacağını göstermiştir.

Gıda işleme firmasının eğer yeni ağızlığı kullanacak olursa ürün satın almayacakların oranının % 10 ' dan az olup olmayacağına karar vermesi gerekmektedir. Tesadüfi olarak seçilen 1000 müşteriden 63 tanesinin yeni durumda peyniri satın almayacakları tespit edilmiştir. Eğer gerçek anakütle oranı % 10 ise gözlemlenen örnek oranının 0,063 ' e eşit veya küçük olma olasılığı nedir (Orhunbilge, 2013)

Örnek 7: Bir üretim sürecinde üretilen 10 diskten birinin hatalı olduğu tespit edilmiştir. Eğer bu süreçten 400 adet örnek tesadüfi olarak seçilirse; hatalı disk oranının 0,10 ile 0,12 arasında olması olasılığı nedir? (Smidt, 2001)

Standart Sapma ile Standart Hatanın Farkı

Standart sapma (standart deviation), bir veri kümesinde değerlerin ortalamaya ne kadar yakın veya uzak olduğunu belirleyen, kareli ortalama şeklinde bir değişkenlik ölçüsüdür. Standart sapmayı anakütle için hesaplayabileceğimiz gibi (σ), örneklem için de hesaplayabiliriz (s).

Diğer yandan, örneklem ortalaması, anakütle ortalaması hakkında elimizde bulunan en iyi bilgidir. Standart hata (standart error) bize, örneklem ortalamasının anakütle ortalamasına ne kadar yakın veya uzak olduğunu gösterir. Satandart hata örnekleme dağılımına, örneklem istatistiğinin dağılımına ilişkindir ve örneklem ortalamalarının oluşturduğu veri kümesinden yararlanılarak verilen formüller ile hesaplanılır (Tablo 2). Anakütle standart sapmasını bilmediğimiz birçok durumda, örneklem standart sapması s' yi kullanarak tahmini bir standart hata değeri olan $s_{\bar{x}}$ değerini kullanabiliriz (Gürsakal, 2013).

	Anakütle İçin	Örneklem İçin
Standart sapma (birimlerin değişkenliği)	$\sigma = \sqrt{\dfrac{\Sigma (x-\mu)^2}{N}}$	$s = \sqrt{\dfrac{\Sigma (x_i - \bar{x}_i)^2}{n-1}}$
Standart hata (ortalamaların değişkenliği)	$\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$	$s_{\bar{X}} = \dfrac{s}{\sqrt{n}}$

Tablo2: : Standart Sapma ve Standart Hata Formülleri (Gürsakal, 2013)

Konu Kaynakları:

Applied Statistics_David M. Levine, Patricia P. Ramsey, Robert K. Smidt_Prentice Hall, 2001

İşletme İktisat için İstatistk_Paul Newbold (Çeviren Ümit Şenesen)_Literatür Yayıncılık,

Sağlık Araştırmaları için Temel İstatistik _ Murat Hayran , Mutlu Hayran _ Omega araştırma, Ankara, Mayıs 2011

İşletmelerde Uygulamalı İstatistik_ İsmail Hakkı Armutlulu _ Alfa, İstanbul, 2008

İstatistik_Burhan Çil_Detay Yayıncılık, Ankara, 2008

Çıkarımsal İstatistik_İstatistik2_Necmi Gürsakal_Dora,Bursa,2013

İşletme İstatistiğinin Temelleri_Bowerman O' Connel (Çeviren Neyran Orhunbilge)_Nobel, Ankara, 2013