

Örneklem Ortalamalarının Dağılımı ve Merkezi Limit Teoremi

Evren: Belirli bir araştırma problemi için araştırmacının çalışmak istediği tüm gruptur.

Örneklem: Evrenin içerisinde seçilen ve onu temsil etmesi amaçlanan birey grubuna örneklem denir.

İstatistik: Örneklemi tanımlayan değer

Parametre: Evreni tanımlayan değer.

Örneklem Hatası: Örneklem istatistiği ile evrendeki karşılığı olan parametre arasındaki farka *örneklem hatası* denir.

Örneklem Dağılımı: Belirli bir örneklem büyüklüğünde (N), seçilebilecek tüm olası örneklemelerden elde edilen istatistiğin dağılımı.

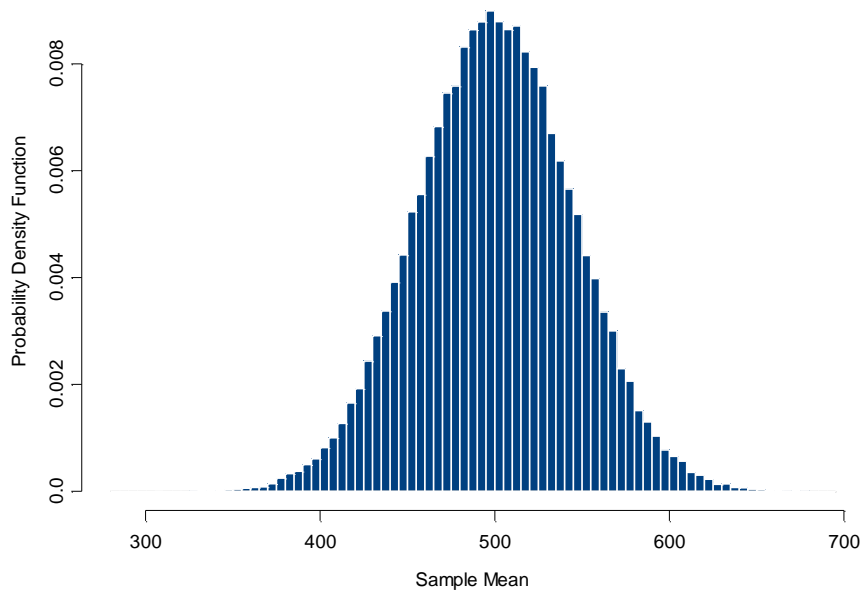
Beklenen Değer: Belirli bir örneklem büyüklüğünde (N), seçilebilecek tüm olası örneklemelerden elde edilen istatistiğin ortalaması evren parametresinin beklenen değerini verir.

Örneklem ortalamalarının dağılımı: Belirli bir örneklem büyüklüğünde (N), seçilebilecek tüm olası örneklemelerden elde edilen ortalamaların dağılımı.

1. Evrenden belirli bir büyüklükte örneklem seçtiğinizi düşünün (Örneğin $N=10$)
2. Örneklemdeki her bir bireyin test puanını elde edelim ($Y_1, Y_2, Y_3, \dots, Y_{10}$)
3. Örneklemdeki test puanlarının ortalamasını alalım ($\bar{Y}_1=520$)
4. 1'den 3'e kadar olan adımları sonsuz kes tekrar ederek örneklem ortalamalarını hesaplayalım ($\bar{Y}_1, \bar{Y}_2, \bar{Y}_3, \bar{Y}_4, \dots, \bar{Y}_\infty$)

$\bar{Y}_1, \bar{Y}_2, \bar{Y}_3, \bar{Y}_4, \dots, \bar{Y}_\infty$ Bu ortalamaların frekans dağılımı bize örneklem ortalamalarının dağılımını verir.

Örn. Evrende ortalaması 500 varyansı 100 olan bir dağılımdan ($\mu = 500, \sigma = 100$) örneklem büyüklüğü 5 olan ($N=5$) 100000 tane örneklem seçilmiş ve elde edilen örneklem ortalaması aşağıdaki grafikte gösterilmektedir.



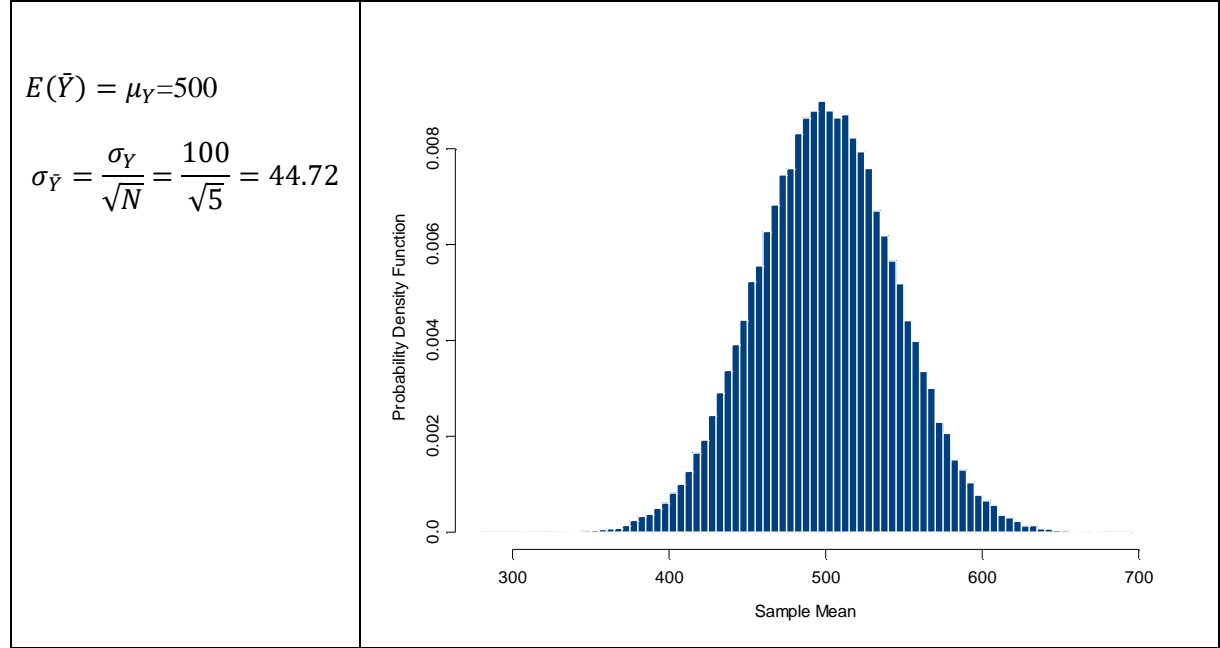
Merkezi Limit Teoremi

Ölçülen özellik evrende normal dağıldığında veya örneklem yeterince büyük olduğunda ($N \geq 30$) Merkezi limit teoremine göre örneklem ortalamalarının dağılımı normal dağılıma yaklaşır. Bu dağılımın ortalaması evren ortalamasının beklenen değeridir. Bu dağılımın standart sapması ise $\frac{\sigma}{\sqrt{N}}$ 'dir.

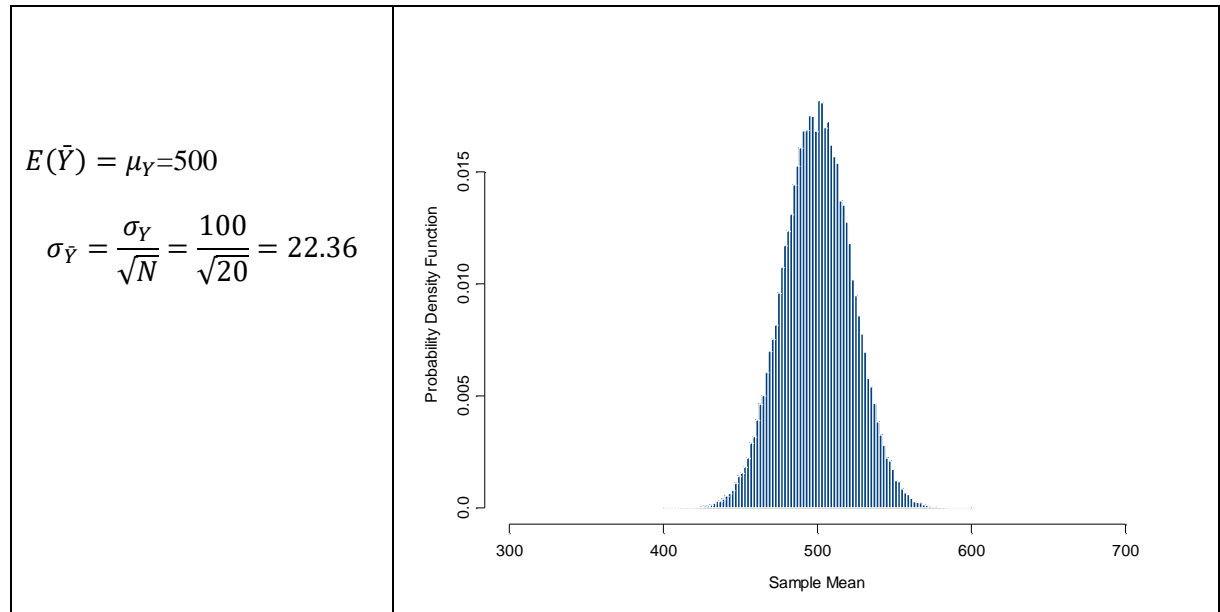
Örnekler

Y değişkeni evrende normal dağılan ortalaması 500 ve standart sapması 100 olan ($\mu_Y = 500, \sigma_Y = 100$) bir değişken olsun:

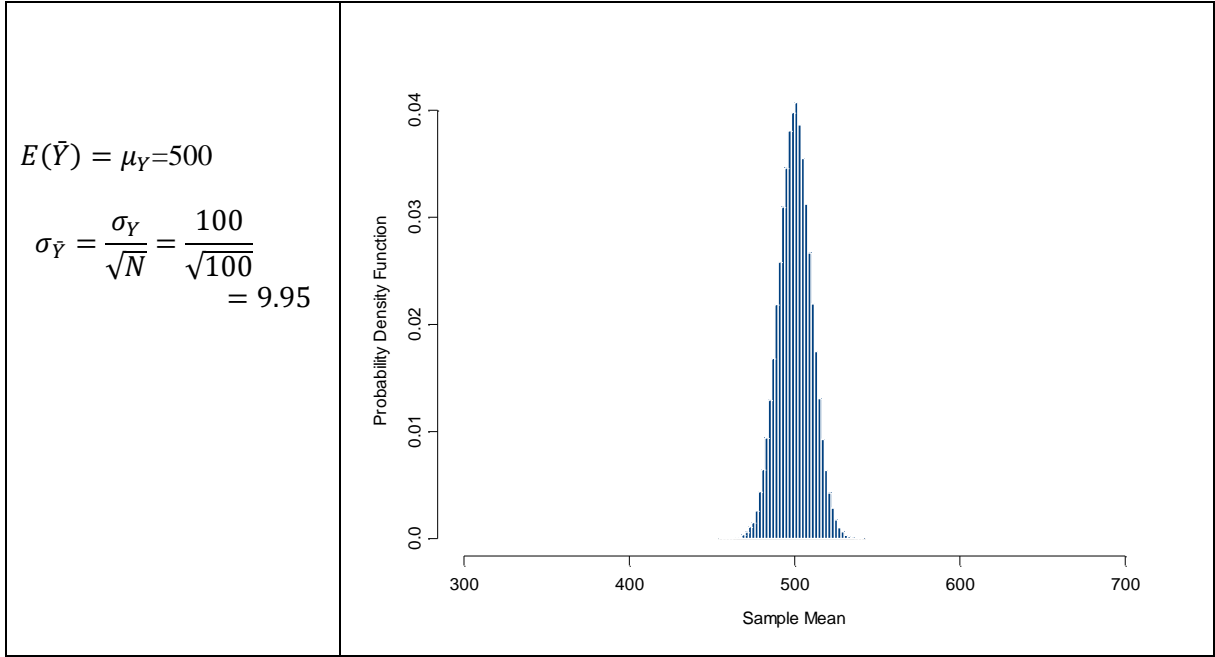
1. $N=5$ için örneklem ortalamalarının dağılımı (100000 örneklem çekildiğinde)



2. $N=20$ için örneklem ortalamalarının dağılımı (100000 örneklem çekildiğinde)



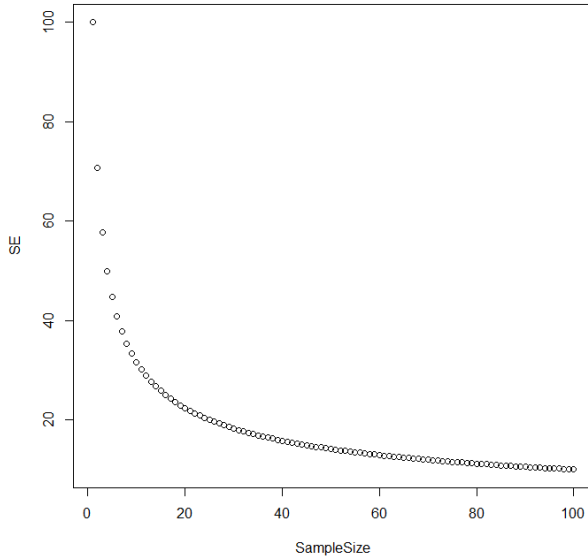
3. N=100 için örneklem ortalamalarının dağılımı (100000 örneklem çekildiğinde)



Büyük sayılar kuralı: Örneklem büyüdükçe örneklem ortalaması evren ortalamasına yaklaşır.

Ortalamanın standart hatası: Örneklem ortalamaları dağılımının standart sapması ortalamanın standart hatasıdır. Standart hata evren ortalaması ile örneklem ortalaması arasındaki beklenen farkın bir ölçümünü verir.

Örneklem büyüdükçe örneklemin standart hatası küçülür.



Olasılık ve Örneklem ortalamalarının dağılımı

Örneklem ortalamalarının dağılımının temel kullanımı evrenden belirli bir örneklemin seçilme olasılığının bulunmasıdır.

Örnek1:

Evrende X testine ilişkin test puanlarının dağılımı normaldir ve ortalaması 500 standart sapması 100'dür ($\mu=500$ ve $\sigma=100$).

Bu evrende rastgele seçilen 25 kişilik bir örneklemin ortalamasının 540'ın üzerinde olması olasılığı nedir ($P(\bar{X}>540)$)?

%2.3

Bu evrenden rastgele seçilen 5 kişilik bir örneklemin ortalamasının 540'ın üzerinde olması olasılığı nedir?

%19

Örnek2:

Evrende ortalaması 40 standart sapması 8 olan bir dağılımdan:

- N=16 kişilik bir örneklem seçiliyor. Seçilen bu örneklemin ortalaması $M=36$ olarak hesaplanıyor. Sizce bu örneklem tipik bir örneklem midir?
- Eğer örneklem büyüklüğü $N=4$ olsaydı bu örneklemini tipik bir örneklem olarak mı düşünürdünüz?

Örnek3:

Evrende ortalaması 530 standart sapması 80 olan bir dağılımdan:

- $N=16$ kişilik bir örneklem seçtiğimizde, örneklem ortalamasının %95 olasılıkla hangi aralıkta olmasını beklersiniz?
- Eğer örneklem büyüklüğü 100 olursa örneklem ortalamasının %95 olasılıkla hangi aralıkta olmasını beklersiniz?

Örnek4:

Evren ortalaması 50 standart sapması 12 olan bir dağılımdan 16 kişilik bir örneklem seçiliyor. Bu örnekleme bir deney uygulanıyor ve deney sonucunda örneklem ortalaması $M=57$ olarak bulunuyor. Sizce Eğer deneyin bireyler üzerinde hiçbir etkisi yoksa bizim evrenden bu örneklemini seçmemiz olası mıdır? Diğer bir ifade ile evrenden 16 kişilik bir örneklem seçtiğimizde örneklem ortalaması % 95 olasılıkla hangi aralıkta olur?

Hipotez Testine Giriş

Örnek4'te aslında bir hipotezi test ediyoruz.

Hipotez testi: Hipotez testi bir evren parametresi hakkında oluşturulan hipotezi örneklem verisi kullanarak test etmeye olanak veren bir istatistik yöntemidir. Gözlem ya da deneme sonucu elde edilmiş sonuçların, *rastlantıya bağlı olup olmadığının* inceler.

İki tür hipotezimiz vardır:

Sıfır(Yokluk) Hipotezi (Null Hypothesis) H_0 : Araştırmalarda test ettiğimiz hipotez genellikle sıfır hipotezi olarak adlandırılır ve aşağıdaki gibi ifade edilir:

$H_0: \mu = 50$ (deneyin uygulanmasına rağmen ortamala hala aynıdır)

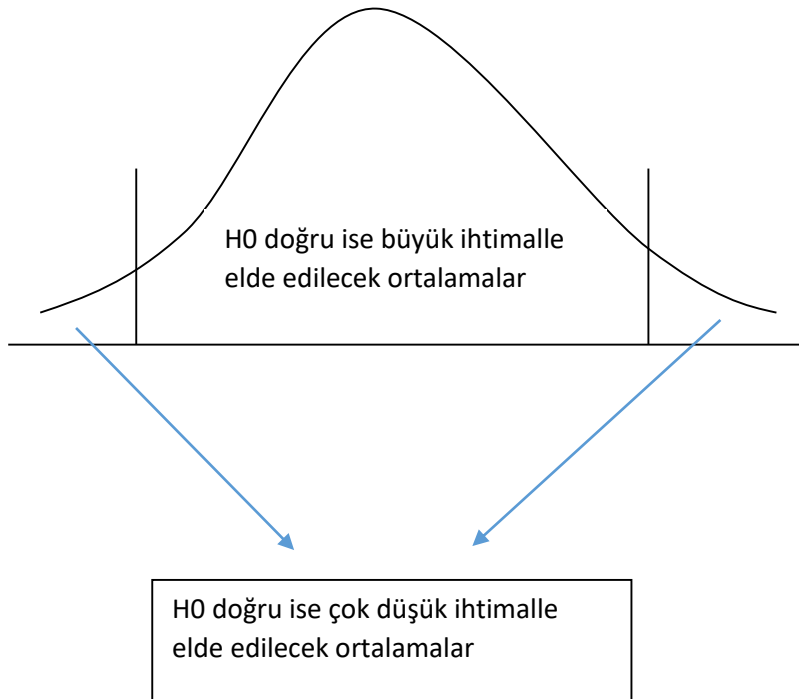
- ✓ Sıfır hipotezi her zaman evrene ilişkin kurulur
- ✓ Çoğu zaman bağımsız değişkenin bağımlı değişken üzerinde bir etkisi olmadığı şeklinde kurulur.
- ✓ Sıfır hipotezi çoğu zaman değişimin, etkinin veya ilişkinin olmadığını ifade edecek şekilde yazılır.

Alternatif (Araştırma) hipotezi H_1 : Sıfır hipotezinin tam tersidir. Değişimi, farkı veya ilişkiyi ifade eder.

$H_1: \mu \neq 50$ (Deneyin etkisiyle artık ortalama 50'den farklıdır)

Red bölgesi/Kritik bölge: Örneklemden elde ettiğimiz istatistiğin bizi H_0 hipotezinin doğruluğunu sorgulattığı değerlerdir.

Karar kuralı: Eğer örneklemden elde ettiğimiz değer kritik bölgedeyse H_0 reddedilir.



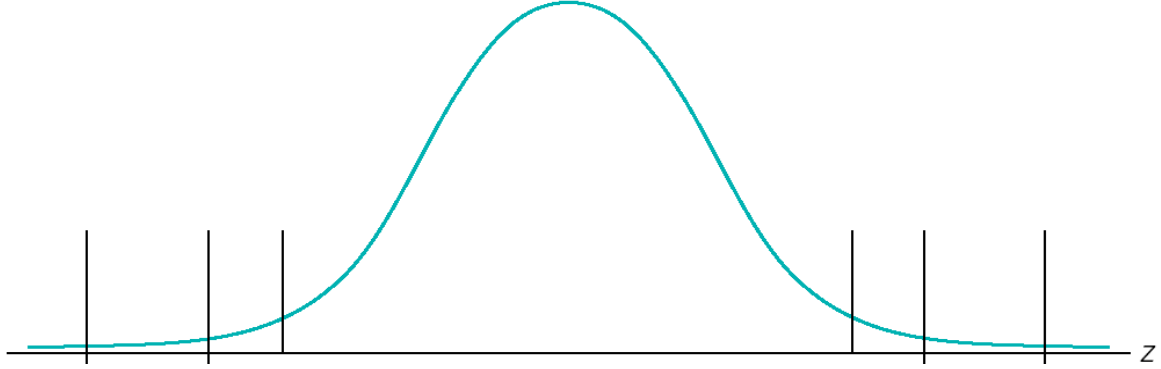
Birinci tip hata (type I error): Doğru olan H_0 hipotezinin reddedilmesiyle ortaya çıkan hatadır. Bazen alfa hata oranı olarak da isimlendirilir.

Kabul edilebilir bir alfa oranı kaçtır?

Önce alfa düzeyini belirleriz bu düzeye göre de kritik bölgeyi/red bölgesini belirleriz.

Alfa düzeyi bize kritik bölgenin oranını/olasılığını verir.

Alfa'nın 0.05, 0.01 ve 0.001'lik değerlerine karşılık gelen z değerlerini hesaplayalım:



Birinci tip hatayı alfa değerini küçültürerek azaltabiliriz.

İkinci tip hata (Type II error): H_0 'ın yanlış olduğu durumlarda H_0 'ı reddedilmemesi durumunda ortaya çıkar.

Bir hipotez testinde olası kararlar:

		H_0	
		Doğru	Yanlış
Karar	H_0 reddedilir	α	
	H_0 kabul edilir		β

- İkinci tip hata H_0 aslında yanlışken örneklem ortalaması kritik bölgede olmadığında ortaya çıkar.
- Deneysel etki küçük olduğunda genelde ortaya çıkar: etkinin büyüklüğü ortalamayı kritik bölgeye atacak kadar büyük değildir.
- İkinci tip hatanın sonuçları genelde birinci tip hatanınki kadar ciddi değildir.
- İkinci tip hata genelde araştırma hipotezlerinin araştırmacının beklediği şekilde çıkmadığını gösterir.
- Tip 1 hatadan farklı olarak tek bir tip 2 hata değer tanımlamak mümkün değildir. Tip II hata oranı birden fazla faktöre bağlıdır.

Örnek: Bir arařtırmacı bitkisel tedavinin bireylerin hafızalarını güçlendirdiğine inanmaktadır. Bu fikrini desteklemek amacıyla $n=16$ kiřiden oluřan bir gruba 90 gün boyunca her gün ginseng çayı içirmiřtir. 90 günün sonunda bu kiřilere bir hafıza testi uygulamıř ve grubun ortalamasını $M=54$ olarak hesaplamıřtır. Standart bir test olan bu hafıza testinin evrende ortalaması 50 ($\mu=50$) ve standart sapması 12'dir ($\sigma=12$).

- a. Arařtırmaya iliřkin arařtırma sorusunu yazınız.
- b. Arařtırmanın hipotezlerini yazınız
- c. Uygun bir alfa düzeyi seçerek red bölgesini tanımlayınız
- d. Bitkisel tedavinin hafıza üzerinde etkisi olup olmadığını yorumlayınız.