Olasılık ve İstatistik

Fırat İsmailoğlu, PhD

Hipotez Testi - I

Karar Vermek

Şimdiye kadar populasyon paramaterlerini (ortalama, populasyon oranı) tahmin etmekle ilgilendik. Fakat bir mühendisin eldeki veriyle yaptığı tek iş bu değildir. Genellikle mühendisten eldeki veri yardımıyla bir karar vermesi de beklenir. Örneğin iki farklı ölçüm cihazı ile yapılan ölçümler arasında bulunan farklar anlamlı mıdır, bir fabrikanın üretittiği mallardan bazıları alınarak oluşturulan örnek, bu fabrikanın ürettiği tüm malları yansıtır mı?

Hipotez Oluşturmak

Bir duruma karar veririken öncelikle bir hipotez oluştururuz; daha sonra elimizdeki verinin hipotezimizi destekleyip desteklemediğine bakarız. Eğer veri hipotezi desteklemezse, hipotezi redederiz, desteklerse hipotezi reddetmeyiz.

Aşağıdaki akış şeması veriye dayanarak karar verme sürecini özetler.

Karar Verme Süreci

ör. Varsayalımki hipotezimiz bir fabrikanın ürettiği malların $\%10$'nun bozuk olduğudur. Bu hipotezi test etmek için fabrikanın ürettiği mallardan rastgele 100 tane seçtiğimizi ve bunların 20 tanesinin bozuk olduğunu varsayalım. Eğer hipotezimiz doğru olasaydı, böyle bir veriyi elde etme olasılığımız çok düşük olurdu. Gerçekten: $p = 0.1$ iken 100 denemede toplam 20 başarının olması olasılığı:

$$P(X = 20) = \binom{100}{20} 0.1^{20} 0.9^{80} = 0.002$$

Bu durumda elimizdeki veri hipotezi desteklemediği için, hipotezimiz reddederiz.

Neyin Hipotez Olacağı

Gördükki elimizdeki veri hipotezi desteklemediğinde (yani hipotezin varlığında elimizdeki gibi bir verinin elde edilme olasılığı çok düşük olduğunda) hipotezi reddediyoruz, veri hipotezi desteklediğinde hipotezi reddetmiyoruz.

Yani karar verme sürecinde aksiyonlarımız yalnızca iki tanedir:

ya: hipotezi reddetmek,

yada : hipotezi reddetmemek.

Başka bir aksiyon mümkün değildir.

Not!!: Hipotezi reddetmemek hipotezi kabul etmek anlamına gelmez; yalnızca eldeki verinin hipotezi desteklemediği anlamına gelir.

Eğer kanıtlamaya çalıştığımız şeyin tersini hipotez olarak düşünürsek, hipotezi reddetme durumunda kanıtlamaya çalıştığımız şeyin tersini reddetmiş oluruz. Yani tersi yanlış olur, böylece kendisi doğru olur.

Bu, aynı Olmayan Ergi Yöntemi gibidir. Kanıtlamak istediğimiz şeyin tersinin yanlış olduğunu, böylece kendinin doğru olduğunu göstermeye çalışıyoruz.

Hipotez Türleri:

1) Sıfır Hipotezi (Null Hypothesis)

Sıfır hipotezi göstermek (kanıtlamak) istediğimiz tezin mantıksal tersidir. Sıfır hipotezi H_0 ile gösterilir.

2) Alternatif Hipotezi (Alternative Hypothesis)

Alternatif hipotez göstermek (kanıtlamak) istediğimiz tezin kendisidir. Sıfır hipoteinin tersidir; H_1 ile gösterilir.

Sonuç olarak sıfır hipotezine, hipotezimiz gibi davranırız. Eğer elimzdeki veri sıfır hipotezini desteklemezse sıfır hipotezi reddedilir, alternatif hipotez kabul edilir. Eldeki veri sıfır hipotezini desteklerse, sıfır hipotezi rededilmez (ama kabul de edilmez).

ör. Diyelimki aşırı kahve tüketiminin kansere yol açtğını göstermek istiyoruz. Bu durumda:

H_0: Aşırı kahve tüketmek kansere neden olmaz.

H_1: Aşırı kahve tüketmek kansere neden olur.

ör. Yeni bir ilaç çıkaran bir ilaç firması, bu ilacın mevcut ilaçdan daha etkili olduğunu göstermek istiyor. Bunun için iki grup insan topluyor: çalışma grubu ve kontrol grubu.

Çalışma grubundakilere yeni ilaç veriliyor; kontrol grubundakilere mevcut olan ilaç veriliyor.

H_0: Kontrol ve çalışma grupları arasında fark yoktur.

H_1: Kontrol ve çalışma grupları arasında fark vardır.

ör. Bir savcı sanığın suçlu olduğuna inanıyor.

H_0: Sanık suçsuzdur.

H_1: Sanık suçludur.

Eldeki veri (bulgu) sanığın suçsuz olduğu ile çelişirse, sanığın suçsuz olduğunu reddeder; sanığın suçlu olduğunu düşürüz; eğer bu veri sanığın suçsuz olduğu ile çelişmezse; sanığın suçsuzluğunu reddedemeyiz, sanığın suçsuz olduğunu çürütecek elimizde yeterince veri yoktur deriz.

Verinin, Hipotezi (H_0)'ı Destekleyip Desteklemediğine Nasıl Karar Vereceğiz?

Diyelimki bir grip aşısı var ve bunun etkisinin 2 yıldan fazla sürme olasılığı 0.25. Bir firma yeni bir grip aşısı çıkarıyor, çıkardıkları bu yeni aşının eski ilaçtan farklı olduğunu düşünüyor. Aşağıdaki şekilde hipotezler kuruluyor:

p, yeni aşının 2 yıldan sonra bile etkili olma olasılığını göstermek üzere:

$H_0 : p = 0.25$ (iki aşı aynıdır)

$H_1 : p \neq 0.25$ (yeni aşı daha iyidir)

(Dikkat edin burda düşündüğümüzün tersini H_0 yaptık)

Firma 20 kişiye kendi aşısını yapıyor. Normalde eğer 20 kişiye eski aşı yapılmış olasydı, $20 \times 0.25 = 5$ kişinin 2 yıldan daha fazla bir süre korunmuş olması beklenirdi. O da bu yüzden kendince şöyle bir karar mekanizmasi kuruyor. Toleransi 2 aliyor, diyorki, 20 kişiden fayda görenlerin sayısı 5-2=3'ten az ise yada 5+2=7'den fazla ise artik veri H_0 hipotezini desteklemiyor olarak düşünebilirim, böylece H_0'ı reddetip H_1'i kabul ederim.

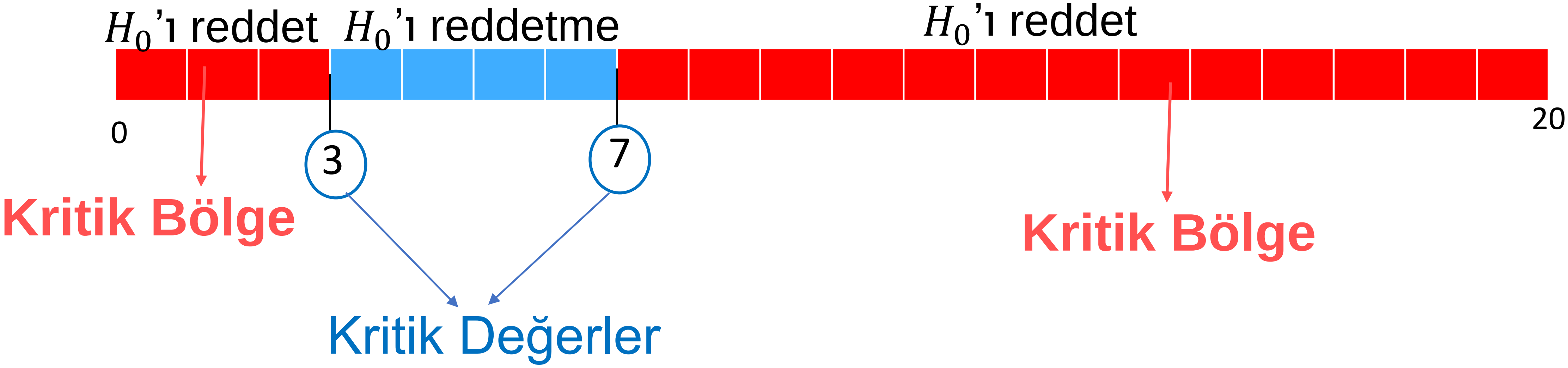
Verinin, Hipotezi (H_0)'ı Destekleyip Desteklemediğine Nasıl Karar Vereceğiz?

Aşı yapılan 20 kişi içersinden, 2 yıl sonra bile yeni aşıdan hala yarar gören kişilerin X binomial rastgele değişkeni ile gösterelim. X, 0 ve 20 arası bir tamsayıdır.

$3 < X < 20$ iken H_0'ı yani iki aşı arasinda bir fark olmadiği hipotezini reddetmeyiz, aksi halde redderiz.

Burada bilmemiz gereken bazı tanımlar var:

1) **Test Istatistiği**: kararımızı (H_0'ın rededilip edilmemesi) dayandırdığımız değişkendir. Yukarıdaki örnekte bu, aşıdan fayda gören kişi sayısıdır.
2) **Kritik Bölge**: Test istatistiği bu bölgede iken H_0 reddedilir (kısaca H_0'ın reddedildigi bölge)
3) **Kritik Değer**: H_0'in reddedilmediği son test istatistikleridir. Bu değerldenb önce yada sonra kritik bölge başlar.



Tip-1 ve Tip-2 Hataları

H_0 reddetme kararını verirken iki tür hata yapabiliriz. Bunlardan ilki H_0 doğru iken H_0'ı reddetme (tip-1 hatası), ikincisi H_0 yanlış iken H_0 ‘ı reddetmeme (tip 2 hatası). Bu hatalar şu şekilde özetlenebilir:

	H_0 doğru	H_0 yanlış
H_0'ı reddet	Tip-1 hatası	Doğru karar ✓
H_0'ı reddetme	Doğru karar ✓	Tip-2 hatası

Şimdi otomatik olarak tolerans değerlerine nasıl karar vereceğimize bakalim. Yani hangi değere kadar H_0'ı reddetmeyeceğizı hangi değerden sonra reddetmeye başlayacağız ona bakalim. Yine başka bir deyişle neye göre ve nasıl kritik değer ve bölgeleri oluşturacağımıza bakalim. Bunun için ilk önce Anlamlılık Düzeyi kavramini bilmemiz gerekir.

Anlamlılık Düzeyi (Level of Significance) (α)

Anlamlılık düzeyi tip-1 hatası yapma olasılığıdır. Yani hipotez doğru iken hipotezi reddetme olasılığıdır. α ile gösterilir.

Önceki örnekte tip-1 hatası yapabilmemiz için hipotez doğru iken (yani $p = 0.25$) iken bunu reddetmemiz gerekir. Reddetme olabilmesi için ise ya 3'ten az kişinin fayda görmesi (yani $X < 3$) yada 7'den fazla kişinin fayda görmesi gerekir (yani $X > 7$). Bu iki durumdan herhangi birinin olma olasılığı

$$\alpha = P(tip - 1\ hatası) = P(X < 3) + P(X > 7)$$
$$= 1 - \big(P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7)\big)$$
$$= 1 - \left(\binom{20}{3} 0.25^3 0.75^{17} + \cdots + \binom{20}{7} 0.25^7 0.75^{13}\right)$$
$$= 0.198$$

Anlamlılık Düzeyi (Level of Significance) (α)

Buna göre, $H_0 : p = 0.25$ olan sıfır hipotezi $\alpha = 0.198$ *anlamlılık düzeyinde ölçülmüştür* denir. Yani toleransi 2 alarak 3'ten az ve 7'den fazla fayda gören kişi olduğunda $H_0 : p = 0.25$ ‘ı reddetmemizle en fazla 0.198 olasılığıyla hata yapmış oluruz.

Tip-1 Hatasını (α'yı) Azaltmak İçin Ne Yapılabilir?

Tip-1 hatasını azaltmak demek; H_0'ı reddetme olasılığını düşürmek demektir. Yani kritik bölgeyi küçültmemiz gerekir (bu bölgede H_0 reddediliyordu). Fakat bunu yaparsak da tip-2 hatasını yapma olasılığımız artar; çünkü reddetmemiz gereken H_0'ları da reddetmemiş oluruz.

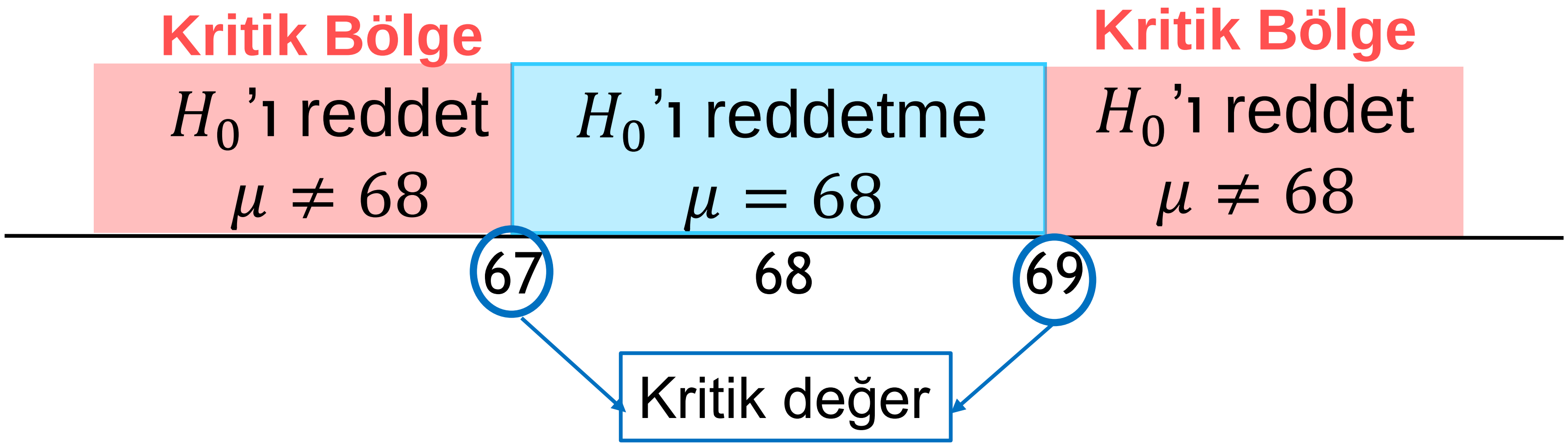
Bu yüzden en güvenli yol, örnek büyüklüğünü (n'i) artırmaktır. Bu şekilde hem tip-1, hem tip-2 hatalarını yapma riskimiz azalır.

ör. Bir okuldaki kişilerin ağırlıklarının ortalaması ile ilgili bir karar vermek istiyoruz. Bu durumda diyelimki sıfır hipotezi ve alternatif hipotez şöyle olsun:

$H_0 \,:\; \mu = 68$ (okuldaki öğrencilerin ağırlıkların ortalaması 68'dir)

$H_1 : \mu \neq 68$ (ağırlık ortalaması 68 değildir.)

Bu durumda okul populasyonundan bir örnek alındığında bu örneğin ağırlık ortalaması 68'e yakın ise H_0 hipotezini reddetmeyiz; yakın değilse H_0 hipotezini reddederiz. Bu yakınlık 68'in ± 1 yanı olsun. Yani örnek ortalaması 69 dan büyük iken yada 67'den küçük iken H_0 reddedilsin.



Şimdi tip-1 hata yapma olasılığını (anlamlılık düzeyini, α'yı) hesaplayalım.

Diyelimki okuldaki kişilerin ağırlıklarının standart sapması $\sigma = 3.6$ olsun ve bu okuldan rastgele 36 kişi seçerek bir örnek oluşturalım.

$\bar{X}$ bu örneğin ortalaması olsun. Merkezi limit teoremi gereği $\bar{X}$ normal dağılıma sahiptir, standart sapması $\frac{\sigma}{\sqrt{n}} = \frac{3.6}{\sqrt{36}} = 0.6$, ortalaması $\mu = 68$ olur.

Şu halde, tip-1 hatası yapabilmemiz için H_0: $\mu = 68$ sıfır hipotezinin doğru olması ama örneğin ortalaması olan $\bar{X}$'in 67 'den az; yada 69'dan fazla olması gerekir (çünkü bu durumlarda sıfır hipotezini reddederiz).

$$\alpha = P(tip - 1\ hatası) = P(\bar{X} < 67) + P(\bar{X} > 69)$$
$$= P\left(Z < \frac{67-68}{0.6}\right) + P\left(Z > \frac{69-68}{0.6}\right)$$
$$= P(Z < -1.67) + P(Z > 1.67)$$
$$= P(Z < -1.67) + 1 - P(Z \leq 1.67)$$
$$= 0.095$$

Okul ortalaması 68 ise 36 büyüklüğündeki bütün örneklerin %9.5'inde, okul ortalamasının 68 olduğunu reddederiz.

Anlamlılık Düzeyini Sabit Tutarak Toleransa Karar Verme

Biraz önce gördüğümüz örneği genelleştirelim.

Diyelimki populasyon ortalaması olan μ'nün bir μ_0 değerine yakın olmadığı hakkında bir karar vermek istiyoruz. Bu durumda sıfır ve alternatif hipotezler:

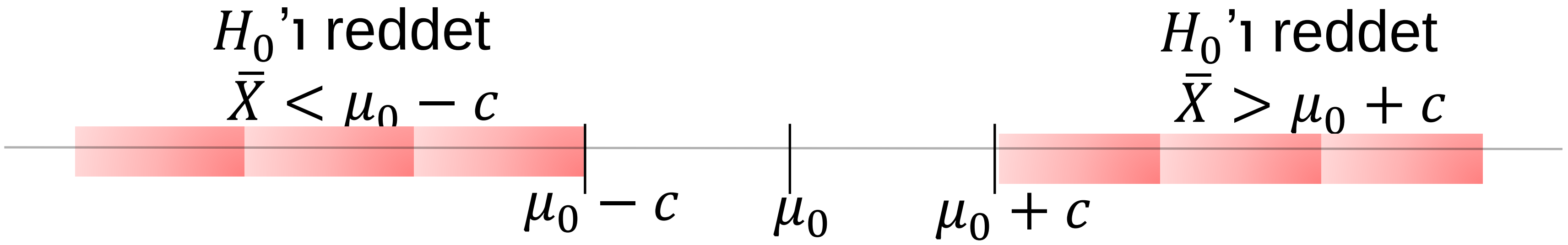
$H_0 : \mu = \mu_0$ ve $H_1 : \mu \neq \mu_0$

H_0'ı reddedip reddetmeyeceğimize karar vermek için c gibi bir tolerans değeri belirleyelim ve *n* büyüklüğünde bir örnek (sample) oluşturalım. Eğer bu örneğin ortalaması olan $\bar{X}$ ile populasyon ortalaması olarak iddia ettiğimiz μ_0 arasındaki fark c 'den büyük ise $H_0 : \mu = \mu_0$ reddedilsin, bu fark c 'den küçükse H_0 reddedilmesin (aralarındaki farkın c'ye kadar olmasına izin veriyoruz, c hariç).

Karar mekanizmamız:

$|\bar{X} - \mu_0| \geq c$ ise $H_0 : \mu = \mu_0$ reddedilsin,

$|\bar{X} - \mu_0| < c$ ise $H_0 : \mu = \mu_0$ reddedilmesin.



Populasyon Varyasyonu Bilinirken Populasyonun Ortalaması Hakkında Karar Vermek

Eğer $\mu = \mu_0$ iken $|\bar{X} - \mu_0| \geq c$ olursa H_0 reddedilir ve tip1-hatası yapılmış olur. Bunun olma olasılığı anlamlılık düzeyi olan α‘ya eşittir:

$\mu = \mu_0$ iken $P(|\bar{X} - \mu_0| \geq c) = \alpha$

Merkezi limit teoremi gereği $\bar{X}$ normal dağılıma sahiptir, merkezi μ_0, standart sapması $\frac{\sigma}{\sqrt{n}}$ dir.

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0)$$

$|\bar{X} - \mu_0| \geq c$ eşitsizliğinde her iki taraf $\frac{\sqrt{n}}{\sigma}$ ile çarpılırsa:

$$\underbrace{\frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_0|} \geq \frac{\sqrt{n}}{\sigma}c$$

$$|Z| \geq \frac{\sqrt{n}}{\sigma}c$$

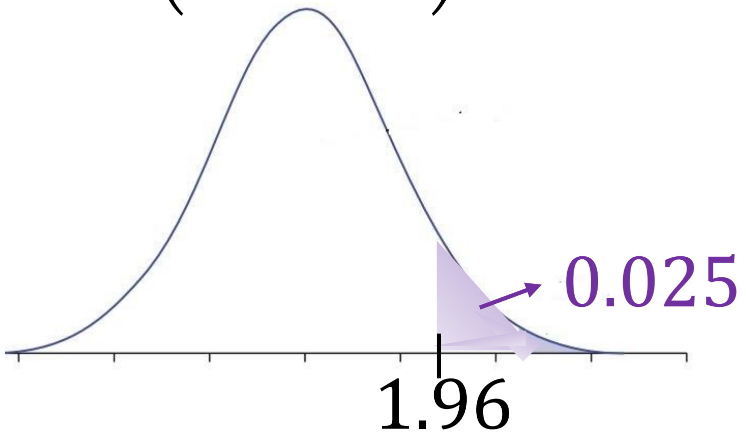
$|\bar{X}-\mu_0| \geq c$ eşitsizliği, $|Z| \geq \frac{\sqrt{n}}{\sigma}c$ eşitsizliğine dönüştürülmüş oldu. O halde

$$P\left(|Z| \geq \frac{\sqrt{n}}{\sigma}c\right) = \alpha$$

Z standart normal rastgele değişkeni merkezi etrafında simetrik olduğundan herhangi bir k değeri için $P(Z \geq k) = P(Z \leq -k)$ eşit olur. Buradan hareketle

$$P\left(|Z| \geq \frac{\sqrt{n}}{\sigma}c\right) = 2 \times P\left(Z \geq \frac{\sqrt{n}}{\sigma}c\right)$$

$$P\left(Z \geq \frac{\sqrt{n}}{\sigma}c\right) = \frac{\alpha}{2}$$



Daha önce gördükki Z standart normal dağılıma sahip rastgele değişkeni için $P(Z \geq 1.96) = 0.025$

$$P\left(Z \geq \frac{\sqrt{n}}{\sigma} c\right) = \frac{\alpha}{2}$$

için, $\frac{\sqrt{n}}{\sigma} c = 1.96$ olursa, $\frac{\alpha}{2} = 0.025$; buradan $\alpha = 0.05$ olur. c yalnız bırakılırsa

$$c = 1.96 \times \frac{\sigma}{\sqrt{n}}$$

olur. Hatrırlarsak, c bizim toleransımız idi. Eğer örnek ortalaması $\bar{X}$ ile populasyon ortalaması olarak düşündüğümüz μ_0 değeri arasındaki fark ($|\bar{X} - \mu_0|$) , c veya , c 'den büyükse $H_0 : \mu = \mu_0$ reddedilsin demiştik.

Şu halde $|\bar{X} - \mu_0| \geq 1.96 \times \frac{\sigma}{\sqrt{n}}$ olursa, $H_0 : \mu = \mu_0$ hipotezi reddedilir. Düzenlersek:

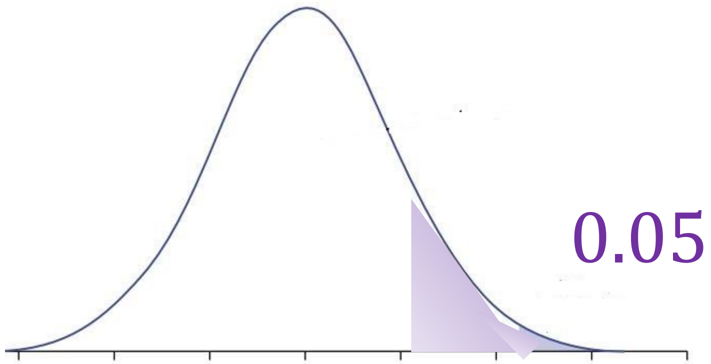
$\frac{\sqrt{n}}{\sigma} |\bar{X} - \mu_0| \geq 1.96$ ise $H_0 : \mu = \mu_0$ reddedilir.

$\frac{\sqrt{n}}{\sigma} |\bar{X} - \mu_0| < 1.96$ $H_0 : \mu = \mu_0$ reddedilmez.

$\frac{\sqrt{n}}{\sigma}|\bar{X}-\mu_0| \geq 1.96$ iken , $H_0 : \mu = \mu_0$ hipotezini reddetmek yönünde yaptığımız kararın yanlış olma olasılığı en fazla 0.05'tir.

$\frac{\sqrt{n}}{\sigma}|\bar{X}-\mu_0|$ değeri test istatistiğimizdir; H_0'ı reddedip reddetmeyeceğimize bu değere göre karar veririz.

Not: Yukarıda 1.96 yerine 1.645 yazılırsa $H_0 : \mu = \mu_0$ hipotezini reddetmekle oluşabilecek hatanın olasılığı 0.1 olur.



ör. Bir önceki örnekteki gibi okuldaki kişilerin ağırlıklarının ortalamasını yine 68 olarak tahmin edelim. Populasyon standart sapması yine $\sigma = 3.6$ olsun ve 36 kişilik bir örnek oluşturalım. Diyelimki bu örneğin ortalaması 71 olsun. Bu durumda $H_0 : \mu = 68$ hipotezini 0.05 anlamlılık düzeyinde reddetmeli miyiz?

Çözüm.

$$\frac{\sqrt{36}}{3.6}|71 - 68| \geq 1.96$$

olduğu için $H_0 : \mu = 68$ hipotezini 0.05 anlamlılık düzeyinde reddederiz. Bu reddetme kararımızın yanlış olma ihtimali en fazla 0.05'tir. Bir başka deyişle %95 eminizki bu kararımız yanlış değildir.