



HACETTEPE
ÜNİVERSİTESİ
İSTATİSTİK BÖLÜMÜ

İST156 İSTATİSTİĞE GİRİŞ II

DERS 7-TESTİN GÜCÜ (POWER OF TEST)

**Ders sorumluları: Prof.Dr.Serpil AKTAŞ ALTUNAY (01 Şubesi)
Doç.Dr. Ayten YİĞİTER (02 Şubesi)**

Hipotez testinde testin gücü ve p-değeri

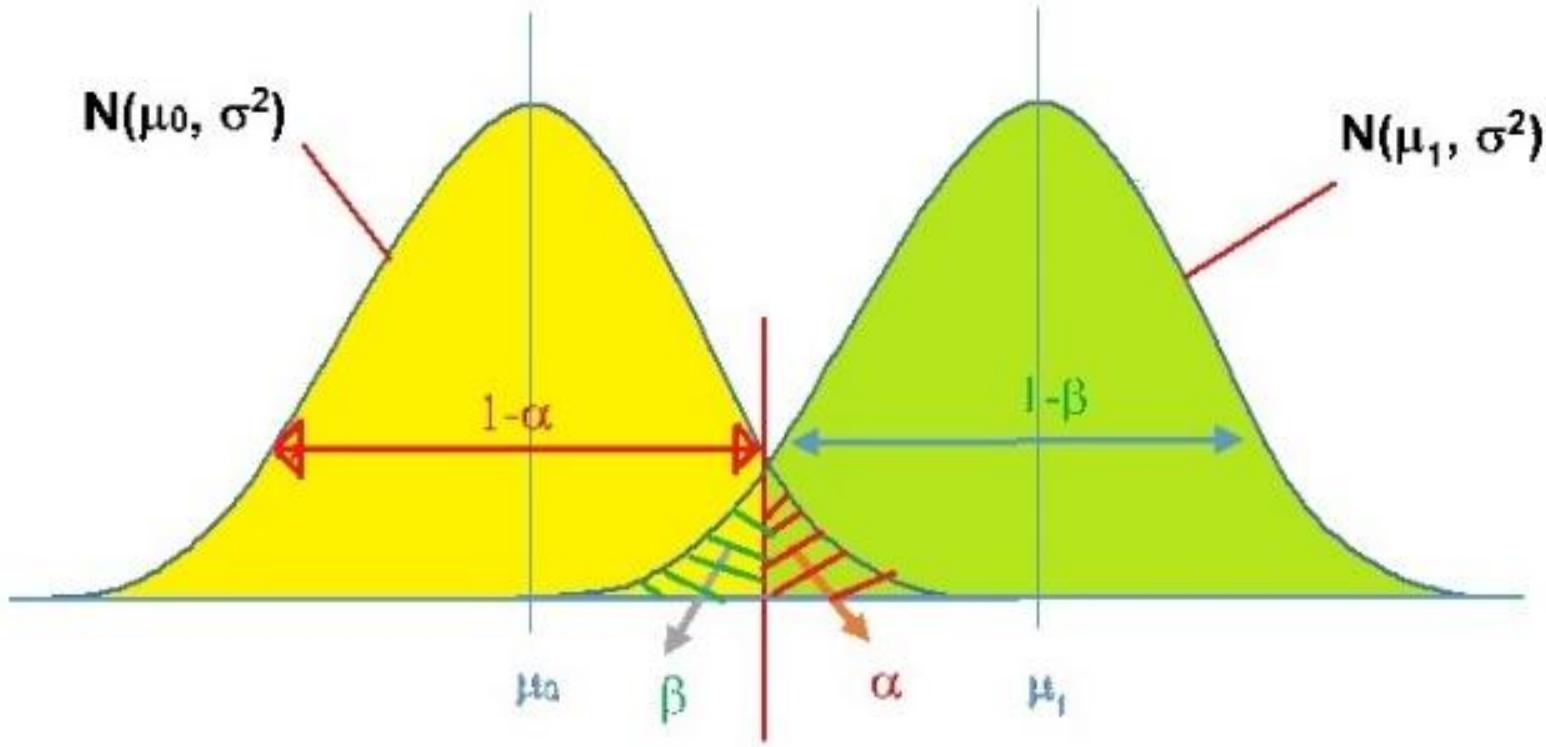
Kitle ortalaması μ 'ye ilişkin hipotez testinde

$H_0: \mu = \mu_0$ hipotezine karşılık

$H_S: \mu > \mu_0$, $H_S: \mu < \mu_0$, $H_S: \mu \neq \mu_0$

seçenek hipotezlerinden birisi kurulur.

Seçenek hipotezleri sırasıyla sağ, sol ya da iki yanlı olarak tanımlanır.



$$H_0: \mu = \mu_0$$

$$H_s: \mu = \mu_1 \quad (\mu_1 > \mu_0) \text{ tek sağ yanlı hipotez}$$

$$\text{Tip I hata } (\alpha) : \quad \alpha = P(H_0 \text{ ret} | H_0 \text{ doğru})$$

$$\text{Tip II hata } (\beta) : \quad \beta = P(H_s \text{ ret} | H_s \text{ doğru})$$

$$\text{Testin gücü: } 1-\beta$$

TESTİN GÜCÜNÜN KAÇ OLMASI BEKLENİR?

TESTİN GÜCÜ :

$P(H_0 \text{ ret} | H_0 \text{ doğru değil})$ olasılığıdır.

TESTİN GÜCÜ $0 < 1 - \beta < 1$ OLDUĞUNDAN ARAŞTIRMACILAR

GENELDE «TESTİN GÜCÜ $\geq 0,80$ » OLMASINI BEKLERLER.

TESTİN GÜCÜ NASIL HESAPLANIR?

Örneğin,

$$H_0: \mu_0 = 170$$

$$H_s: \mu_1 > 170$$

hipotezi ve $\bar{X} = 172$ olsun. $\mu_1 > 173$ için α , β ve $1 - \beta$ şöyle tanımlanır.

$$\alpha = P(\bar{X} \geq 172 | \mu_0 = 170)$$

$$\beta = P(\bar{X} < 172 | \mu_1 = 173)$$

$$1 - \beta = P(\bar{X} \geq 172 | \mu_1 = 173)$$

Örnek: α ve β 'nın hesaplanması

Yeni bir ilaçla ilgili yokluk hipotezi $\theta = 0.90$ 'ı seçenek hipotez $\theta = 0.60$ 'a karşı test edilmek istendığını varsayalım. Test istatistiği X , 20 denemedede (deneysel birimler veya hastalar) gözlenen başarı (iyileşme) sayısıdır ve $x > 14$ ise sıfır hipotezini kabul edecektir; aksi takdirde reddedecektir. α ve β 'yi bulunuz.

Bu örnekte, yokluk hipotezinin kabul bölgesi $x=15,16,17,18,19, 20$, ve ret bölgesi (ya da kritik alan) $x=0,1,2,\dots,14$.

$$\alpha = P(X \leq 14 | \theta = 0.90) = \sum_{x=0}^{14} \binom{20}{x} (0.90)^x (0.10)^{20-x} = 0.0114$$

$$\beta = P(X > 14 | \theta = 0.60) = \sum_{x=15}^{20} \binom{20}{x} (0.60)^x (0.40)^{20-x} = 0.1255$$

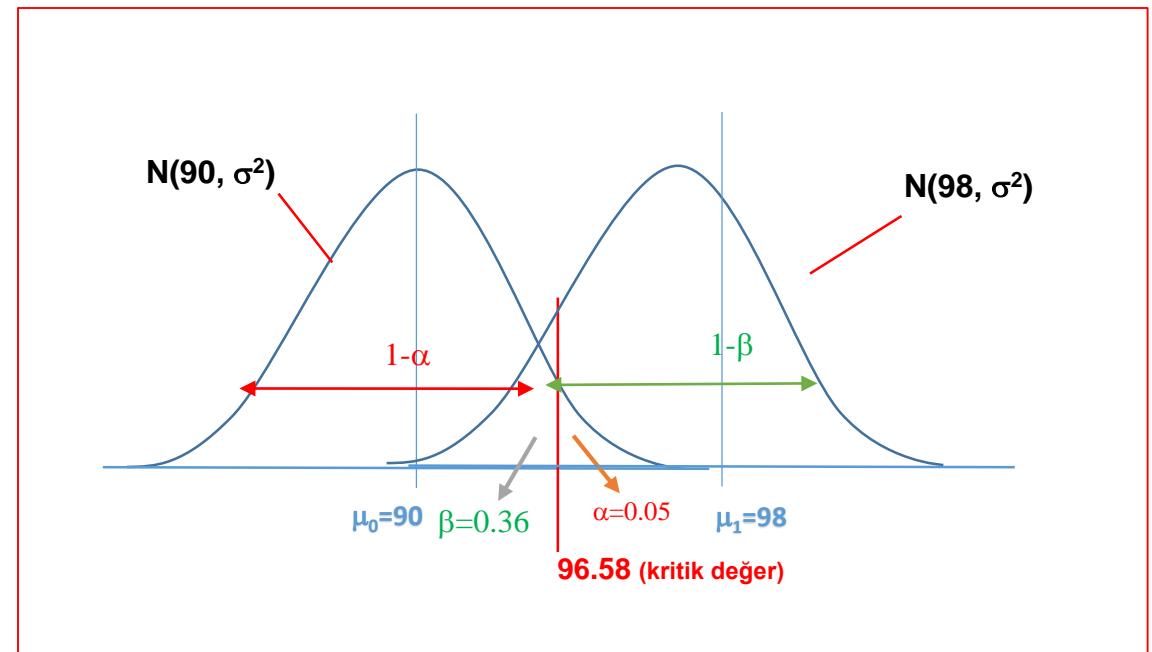
İyi bir test prosedürü, hem α hem de α ve β 'nın küçük olduğu ve bu sayede doğru kararı alma şansımızın yüksek olduğu bir prosedürdür. Bu örnekte Tip II hatanın olasılığı oldukça yüksektir, ancak bu, kritik bölgeyi uygun şekilde değiştirerek azaltılabilir. Örneğin, bu örnekte kabul bölgesi $x > 15$ 'i kullanırsak ve kritik bölge $x \leq 15$ olursa, bunun $\alpha = 0,0433$ ve $\beta = 0,0509$ olacağı kolayca kontrol edilebilir. Görüldüğü gibi, n sabit tutulduğu sürece, bir tip hatanın olasılığı azalırsa, diğer tip hatanın olasılığı artar. Her iki tip hatanın olasılığını azaltmanın tek yolu, örneklem boyutunu artırmaktır.

Örnek 1: Türkiye'deki yetişkinlerin IQ ortalamasının 90 olduğu iddia edilmektedir. Yetişkinlerin IQ ölçümlerinin standart sapması 16 olan normal dağılıma uyduğu varsayılsın. Rastgele seçilen 16 kişi üzerinden $H_0: \mu_0=90$ karşı $H_s: \mu_1>90$ hipotezinin testini $\alpha=0,05$ anlamlılık düzeyinde test ediniz. Ortalama IQ'nun $\mu_1=98$ ($\mu_1>90$) olduğu durumda, testin gücü ne olur hesaplayınız.

$$H_0: \mu_0=90 \quad z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{\bar{X}-90}{16/4} \geq \overbrace{z_{\alpha=0.05}}^{kritik\ değer} = 1.645 \text{ ise } H_0 \text{ hipotezi reddedilir.}$$

$$H_s: \mu_1>90 \quad \frac{\bar{X}-90}{4} = 1.645 \rightarrow \bar{X} = 90 + 1.645 \times 4 = 96.58$$

$$\begin{aligned} \text{Güç} &= P(\bar{X} > 96.58 \mid \underbrace{\mu_1 = 98}_{H_s: \text{doğru}}) \\ &= P\left(Z > \frac{96.58 - 98}{\frac{16}{4}} \mid \underbrace{\mu_1 = 98}_{H_s: \text{doğru}}\right) \\ &= P(Z > -0.355 \mid \mu_1 = 98) \\ 1 - P(Z < -0.355) &= 1 - 0.3613 \\ \text{Güç}(1 - \beta) &\cong 0.6387 \text{ testin gücü} \end{aligned}$$



Örnek 2: Hektar başına düşen buğday miktarı standart sapması 6 olan normal dağılıma uymaktadır. Ziraatçılar hektar başına düşen buğday miktarını 40 kiloya çıkarmak istemektedirler. $\alpha=0,05$ anlamlılık düzeyinde $H_0: \mu_0=40$ 'a karşı $H_s: \mu_1=45$ ($\mu_1>40$) hipotezinin testi için testin gücünün 0,95 olabilmesi için gerekli örneklem büyüklüğünü bulunuz.

$$H_0: \mu_0=40$$

$$H_s: \mu_1=45 \quad (\mu_1>40)$$

$$z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{\bar{X}-40}{6/\sqrt{n}} \geq \overbrace{z_{\alpha=0.05} = 1.645}^{\text{kritik değer}} \text{ ise } H_0 \text{ hipotezi}$$

$$\frac{\bar{X}-40}{6/\sqrt{n}} = 1.645 \rightarrow \bar{X} = 40 + 1.645 \times \frac{6}{\sqrt{n}}$$

$$\beta = P\left(\bar{X} < 40 + 1.645 \times \frac{6}{\sqrt{n}} \mid \underbrace{\mu_1 = 45}_{H_s: \text{doğru}}\right) = 0.05$$

$$P\left(Z < \frac{40 + 1.645 \times \frac{6}{\sqrt{n}} - 45}{\frac{6}{\sqrt{n}}} \mid \underbrace{\mu_1 = 45}_{H_s: \text{doğru}}\right) = 0.05$$

$$\frac{40 + 1.645 \times \frac{6}{\sqrt{n}} - 45}{\frac{6}{\sqrt{n}}} = -1.645$$

$$2 \times 1.645 \times \frac{6}{\sqrt{n}} = 5 \rightarrow n = \left(2 \times 1.645 \times \frac{6}{5}\right)^2 = 15.5867 \cong 16$$

Örnek 3: Yapılacak seçimde bir partinin adayının tercih edilmesi oranı p olmak üzere, bu oran için 1000 seçmenle görüşülmüştür. $\alpha=0,05$ anlamlılık düzeyinde $H_0: P_0=0.5$ karşı $H_s: P_1<0.5$ ($P=0.48$) hipotezinin testi için testin gücünü hesaplayınız.

$$H_0: P_0=0.5 \quad z = \frac{\hat{p}-P_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\hat{p}-0.5}{\sqrt{\frac{0.5 \times 0.5}{1000}}} < \overbrace{z_{\alpha=0.05} = -1.645}^{\text{kritik değer}} \text{ ise } H_0 \text{ hipotezi reddedilir.}$$

$$H_s: P_1<0.48 \quad \frac{\hat{p}-0.5}{0.016} = -1.645 \rightarrow \hat{p} = 0.5 - 1.645 \times 0.016 = 0.47$$

$$\begin{aligned} \text{Güç} &= P\left(\hat{P} < 0.47 \mid \overbrace{P_1 = 0.48}^{H_s:\text{doğru}}\right) \\ &= P\left(Z < \frac{0.47 - 0.48}{\sqrt{\frac{0.48 \times 0.52}{1000}}} \mid \overbrace{P_1 = 0.48}^{H_s:\text{doğru}}\right) \end{aligned}$$

$$\begin{aligned} &= P(Z < -0.379 \mid P_1 = 0.48) \\ &= 0.3556 \end{aligned}$$

$1 - \beta \cong \%36$ testin gücü

Bu testin gücü yaklaşık %35.2'dir. Yani, gerçek oran 0.48 olduğunda bu testin H_0 hipotezini reddetme olasılığı %35.2'dir.

Örnek 4: Bir metalin sertliğinin kalitesinin incelenmesi malzeme bilimde önemli bir yer tutar. İnşaatta kullanılan demirin sertliği, standart sapması 10 kg/mm^2 ile normal dağılmaktadır. Rastgele olarak seçilen 25 birimlik demir çubuk alınmakta; demir çubuklarının ortalama sertliği için $H_0: \mu_0 = 170$ karşı $H_s: \mu_1 > 170$ hipotezi test edilmek istenmektedir. Metalürji mühendisi, örneklem ortalaması 172 ve daha yukarı olduğunda H_0 hipotezini ret edecektir. Buna göre I. Tip hatanın büyüklüğünü hesaplayınız.

$$H_0: \mu_0 = 170 \quad z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{X} - 170}{10/5}$$

$$H_s: \mu_1 > 170 \quad \frac{\bar{X} - 170}{10/5} \geq z_\alpha \text{ ise } H_0 \text{ hipotezi reddedilir.}$$

$$\begin{aligned} \alpha &= P(\bar{X} \geq 172 \mid \underbrace{\mu_0 = 170}_{H_0: \text{doğru}}) \\ &= P\left(Z \geq \frac{172 - 170}{\frac{10}{5}} \mid \underbrace{\mu_0 = 170}_{H_0: \text{doğru}}\right) \\ &= P(Z \geq 1 \mid \mu_0 = 170) \\ &= 0.5 - \overbrace{P(0 < Z < 1)}^{0.3413} = 0.1587 \cong 0.16 \end{aligned}$$

I.TİP VE II.TİP HATA TÜRLERİNİN MAKİNE ÖĞRENMESİNDE KULLANIMI

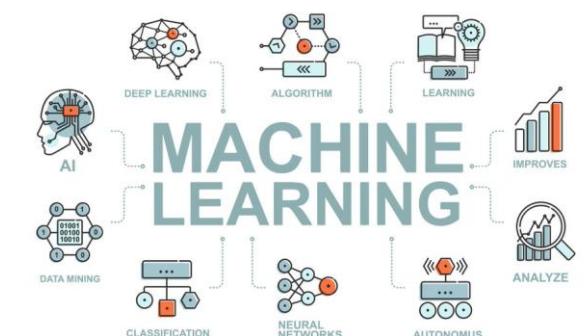
I.Tip ve II.Tip Hata

Kitledeki Durum		
Test Sonucu	H_0 Doğru	H_0 Yanlış
H_0 Kabul	Doğru Karar	II.Tip Hata (β)
H_0 Ret	I.Tip Hata (α)	Doğru Karar

β : II.Tip Hata

$1-\beta$: Testin gücü (Power of test)

α : I.Tip hata



Bu hipotezler için olası dört durumu inceleyelim.

Örneğin bir e-posta koruma sistemi yazılıyor ve sistemdeki izinsiz gönderileri (spam mail) tespit etmek istiyoruz. Yazılımımız bu gönderileri tespit edip engelleyecek.

H_0 : Gelen posta spam değildir.

H_s : Gelen posta spamdır.

Durum 1: Gelen posta spam değil ve bizim yazılım o postaya normal dediyse.....**DOĞRU KARAR**

Durum 2: Gelen posta spam değil ama bizim yazılım onu spam'a atıyorsa.....**I.TİP HATA**

Durum 3: Gelen posta spam ve ve yazılım onu spam'a atıyorsa.....**DOĞRU KARAR**

Durum 4: Gelen posta spam ancak yazılım onu normal postaya atıyorsa.....**II.TİP HATA**

		TAHMİNLER	
GERÇEKTE OLANLAR		Mail spam'dır (1)	Mail spam değildir (0)
Mail spam'dır (1)	True Positive (TP)	False Negative (FN)	
	False Positive (FP)	True Negative (TN)	

KARMAŞIKLIK MATRİSİ (CONFUSION MATRIX)

		Tahmin (Prediction)	
Gerçek (Actual)		Pozitif	Negatif
Pozitif		Doğru Pozitif (DP) (TP)	Yanlış Negatif (YN) (FN)- β
Negatif		Yanlış Pozitif (YP) (FP)- α	Doğru Negatif (DN) (TN)

1. Doğruya doğru demek (True Positive – TP) DOĞRU
2. Yanlışa yanlış demek (True Negative – TN) DOĞRU
3. Doğruya yanlış demek (False Positive – FP) YANLIŞ
4. Yanlışa doğru demek (False Negative – FN) YANLIŞ

YP: I.TİP HATA 

YN : II.TİP HATA 

Test Sonucu	Gerçek Durum	
	Covid-19 (+)	Covid-19 değil (-)
H_0 ret (Hasta) (+)	Doğru (+)	Yanlış (+) Tip I Hata (α)
H_0 reddedilmez (Sağlıklı) (-)	Yanlış (-) Tip II hata (β)	Doğru (-)

$$\text{specificity} = \frac{\text{doğru}(-)}{\text{yanlış}(+) + \text{doğru}(-)} = \frac{\text{doğru}(-)}{\text{gerçek} (-)}$$

$$\text{sensitivity} = \frac{\text{doğru}(+)}{\text{doğru}(+) + \text{yanlış}(-)} = \frac{\text{doğru}(+)}{\text{gerçek} (+)}$$

True positive: Hastaya hasta demek

False positive: Sağlıklıya hasta demek

True negative: Sağlıklıya sağlıklı demek

False negative: Hastaya sağlıklı demek

https://en.wikipedia.org/wiki/Sensitivity_and_specificity

PREDICTIVE VALUES

		POSITIVE (CAT)	NEGATIVE (DOG)
ACTUAL VALUES	POSITIVE (CAT)	TRUE POSITIVE 3  YOU ARE A CAT	FALSE NEGATIVE 1  TYPE II ERROR
	NEGATIVE (DOG)	FALSE POSITIVE 2  TYPE I ERROR	TRUE NEGATIVE 4  YOU ARE NOT A CAT

SINIFLAMADA KULLANILAN BAZI METRİKLER

$$\text{SENSITIVITY (RECALL) (DUYARLILIK)} = \frac{DP}{DP + YN}$$

$$\text{SPECIFICITY (ÖZGÜLLÜK)} = \frac{DN}{DN + YP}$$

$$\text{ACCURACY (DOĞRULUK)} = \frac{DP + DN}{DP + YP + DN + YN}$$

Örnek 5: Bir makine öğrenmesi algoritması yardımcı ile 100 tane fotoğraf, «kedi» ve «kedi değil» olarak sınıflandırılmıştır.

		Tahmin
Gerçek	Kedi(+)	Kedi değil(-)
Kedi(+)	42 (DP)	15 (YN)
Kedi değil(-)	9 (YP)	34 (DN)



$$SENSITIVITY = \frac{DP}{DP + YN} = \frac{42}{42 + 15} = 0,736 = \%74$$

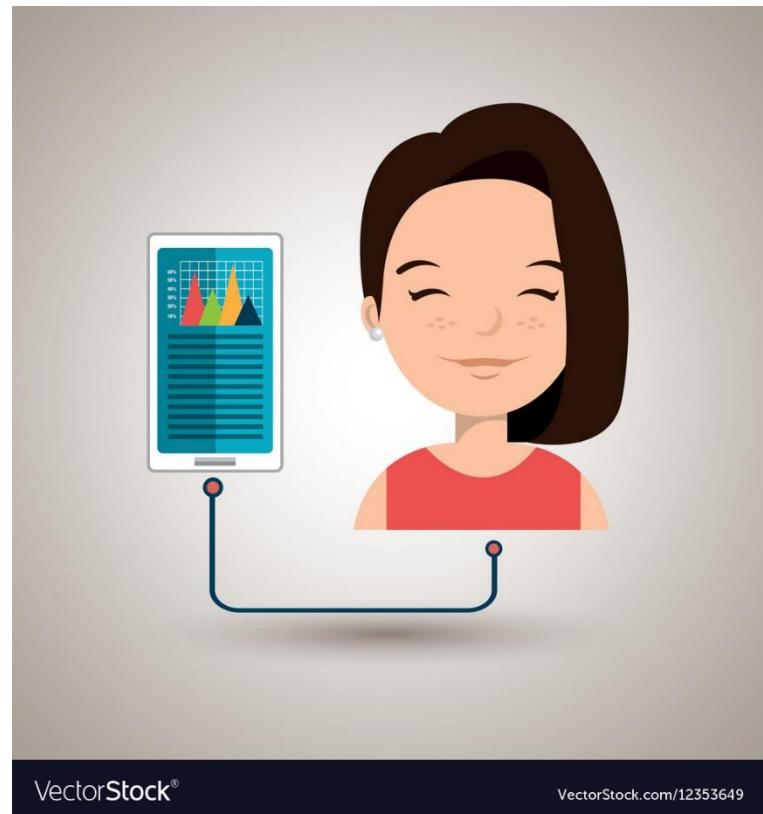
Gerçek kedi fotoğrafları içinde algoritmanın kedileri doğru tespit etme oranı %74'dür.

$$SPECIFICITY = \frac{DN}{DN + YP} = \frac{34}{34 + 9} = 0,790 = \%79$$

Gerçekte kedi fotoğrafı olmayanlar içinde algoritmanın kedi olmayanları doğru tespit etme oranı %79'dur.

$$ACCURACY = \frac{DP + DN}{DP + YP + DN + YN} = \frac{42 + 34}{42 + 15 + 9 + 34} = 0,76 = \%76$$

Gerçekte kedi olan ve olmayan fotoğrafları, tüm durumlar içinde algoritmanın doğru tespit etme oranı %76'dır.



Bir sonraki derste kitle örneklem büyüklüğü konusu işlenecek.



KAYNAK

«GENÇ AKADEMİSYENLER İÇİN BİLİMSEL ARAŞTIRMA TEKNİKLERİ VE YAYIN ETİĞİ» Nobel Yayınevi, 2022.

BÖLÜM 4: ÖRNEKLEM SEÇİMİ VE GÜC ANALİZİ

Hande Konşuk Ünlü - Hacettepe Üniversitesi Halk Sağlığı Enstitüsü

Serpil Aktaş Altunay - Hacettepe Üniversitesi Fen Fakültesi İstatistik Bölümü

Mustafa Altındış - Sakarya Üniversitesi Tıp Fakültesi Tibbi Mikrobiyoloji AD

<https://online.stat.psu.edu/stat415/book/export/html/845>